Modeling growth responses to mid-rotation silvicultural treatments for loblolly pine (*Pinus taeda* L.) plantations in the southeastern U.S.

by

Mauricio Zapata Cuartas

(Under the Direction of Bronson P. Bullock)

Abstract

Loblolly pine (*Pinus taeda* L.) plantations are extensive in the southeastern United States and represent a significant component of the forest products market in this region. For optimal stand-level management decisions, the growth response to any combination of mid-rotation silvicultural treatments like fertilization, thinning, or competing vegetation management needs to be accurately predicted over the long-term. This dissertation presents a review of the most common mid-rotation silviculture treatments applied in loblolly pine plantations and provides a conceptual framework about their effect on growth. The Growth and Yield system presented here consists of a novel taper equation based on a penalized spline regression, a compatible dynamic growth system of differential equations for dominant height, basal area, and stand density, which includes a growth modifier to account for mid-rotation silvicultural effects, and a novel method to recover the diameter distribution for projected stand after mid-rotation silvicultural treatments. The Growth and Yield system and the recovery diameter model were fitted with the regional Thinning and Mid-Rotation Treatment (MRT) study established by the Plantation Management Research Cooperative

(PRMC) at the University of Georgia, which includes various site qualities and growth conditions in thinned and non thinned stands. Therefore, the proposed model well represents regional growth conditions. Additionally, the models presented are new approaches for the southern loblolly pine plantation modeling and were shown to be improvements over existing practices.

INDEX WORDS: Biometrics, Forest growth and yield modeling, Silvicultural treatment responses.

Modeling growth responses to mid-rotation silvicultural treatments for loblolly pine (*Pinus taeda* L.) plantations in the southeastern U.S.

by

Mauricio Zapata Cuartas

B.S.F.E., The National University of Colombia, Colombia, 2001

A Dissertation Submitted to the Graduate Faculty of the
University of Georgia in Partial Fulfillment of the Requirements for the Degree.

Doctor of Philosophy

Athens, Georgia

2020

Modeling growth responses to mid-rotation silvicultural treatments for loblolly pine (*Pinus taeda* L.) plantations in the southeastern U.S.

by

Mauricio Zapata Cuartas

Major Professor: Bronson P. Bullock

Committee: Cristian R. Montes
Michael Kane
Nicole Lazar

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
December 2020

DEDICATION

I would like to dedicate this dissertation to my wife Angela and my two daughters Emilyn and Valeria. They have always been such an incredible source of inspiration, love, and support. To the women whose blood I carry – my mother Luz Marina and my maternal grandmother, Rosa Antonia Lopera (1928-2020).

ACKNOWLEDGMENTS

I am incredibly grateful to my major professor, Dr. Bullock, who guided and supported me through this entire process and recognized my academic strength when I began my journey as a doctoral student, and pushed me beyond. I am forever grateful to you. I am also thankful to Dr. Montes, I will always cherish your warmth and attention to the details.

I am very grateful to the Warnell School of Forestry and Natural Resources for the economic support and the assistantship offered during my doctoral studies. A special thanks to my committee members Dr. Bullock, Dr. Montes, Dr. Kane, and Dr. Lazar, for their invaluable insight, suggestions, and contributions to this work. I want to acknowledge the Plantation Management Research Cooperative (PMRC) members for finance the establishment, maintenance, and data collection of the Mid-Rotation Treatment Experiment. The PMRC for providing the data and support to accomplish my research objectives.

To the PMRC grad students. Thank you for the constant laughs, camaraderie, and for cheering me on through this process.

I questioned whether this day would ever arrive, so I thank you, Dr. Del La Torre and Ms. Constansa Beron. Your frank and practical advice help my family and me during this journey. I am also eternally grateful to your family to open their hearts and home and adopt me as one of their own. Special thanks to Dr. Restrepo and Ms. Adriana M. Rodriguez for belief in my capabilities, encouraged me through my studies, and offer a hand unconditionally when I needed it.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Literature Review

1.1 Introduction

Pine plantations have made a significant expansion across the southeastern United States. Before 1952, pine plantations in this region were marginal; during 1952-2010, the planted area accumulated 39 million acres (19 percent of the total forest area in the Southeast) (Wear and Greis, 2012). Forecasts reveal a positive rate of change of conversion from natural regenerated pine types to pine plantations. By 2060, plantations are expected to represent between 24 and 36 percent of the South's forest area. (Huggett et al., 2013). That means a net increment between 7.8 to 28.2 million acres. Consequently, with this expansion, the southern timber production represents about 15.8 percent of global wood production (Prestemon and Abt, 2002), turning these plantations into the principal source of timber and forest products in the United States.

Since 1960 this pine plantation expansion has been complemented with improvements in silvicultural practices and intensive management, which have shown a remarkable produc-

tivity gain (Fox et al., 2004). Loblolly pine (*Pinus taeda* L.) is the more planted species throughout the Southeast and managed with intensive silviculture practices.

Silvicultural practices serve to increase productivity in loblolly pine plantations. For example, early control of competing woody and herbaceous vegetation increases wood volume by 23-121 percent when evaluated at age 15 (Miller et al., 2003). Mid-rotation fertilization with nitrogen (N; 150 to 200 pounds per acre) and phosphorous (P; 25 to 50 pounds per acre) on a majority of soil types throughout the southern U.S. produces a large and consistent growth response in volume of approximately 25 percent (Fox et al., 2007). N and P fertilization with responses of approximately 55 ft^3 ac^{-1} yr^{-1} have shown to enhance the profitability of loblolly pine plantations (Fox et al., 2007). Thinning is another important silvicultural practice in loblolly pine plantations. Thinning not only changes the structure of the stand (diameter distributions, and changes in spatial competition) but alters the post-thinning height growth of dominant and co-dominant trees (Sharma et al., 2006), and improve the stem quality of the residual stand. Mid-rotation silvicultural practices have been used extensively in the Southern to raise the quality and proportion of timber products in loblolly pine plantations, motivated by the trend in price differentiation from timber products. Forest researches have measured the effect of silvicultural treatments thorough field trials and repeated measurement plot experiments. However, this method is challenging to replicate and cannot be installed elsewhere due to its expensive and time-consuming nature (Weiskittel, 2011). Therefore, investors and forest managers need accurate growth and yield models (G&Y) suitable to predict the expected plantation responses after mid-rotation practice and make decisions. This dissertation proposes a new class of G&Y models for loblolly pine, considering mid-rotation silvicultural practices. Novel empirical approaches for modeling tree and stand attributes are presented. At the tree-level, a new taper equation for loblolly pine is proposed using semiparametric regression. A comparison with traditional parametric taper equations indicates that the new approach enhances the accuracy of stem

diameter and merchantable volume estimations. At the stand-level, a new dynamic system of equations is proposed to project the main stand attributes after mid-rotation silvicultural practices. The system considers the following mid-rotation scenarios: No-intervention, thinning only, thinning and fertilization, thinning and competitive vegetation control, and finally, the combination of thinning, fertilization, and competitive vegetation control. For the tree-level taper equation, this research uses a stem analysis data set for 147 loblolly pine trees extracted from non-thinned planted stands in the Whitehall Forest at the University of Georgia during the fall seasons from 2014 to 2018. Undergraduate and graduate students performed the labor during field practices in the Forest Mensuration course at the Warnell School of Forestry and Natural Resources. A description of this dataset is in Chapter 3. The Thinning and Mid-Rotation Treatment Study (MRT) established by the Plantation Management Research Cooperative (PMRC) at the University of Georgia, was used to develop the stand-level G&Y model system. The MRT is a regional study with 49 sample installations distributed across nine U.S. Southern states with 245 measurement plots. Chapter 4 provides a complete description of the MRT study. Finally a new procedure for predicting future stand diameter distribution given the main stand-level attributes projections and the initial diameter distribution was developed based on copulas of survival diameter distributions at two point times. These new method can be used in combination with a G&Y model system to evaluate the effect of mid-rotation treatments on the availability of wood products at rotation age.

1.2 Modeling Growth and Silvicultural Treatment Effects

Here, treatment effects refer to the additional growth at the stand-level (or plot-level) due to the treatment. Under a regular experiment setup, the condition without treatment serves

as a reference (also called the control case) to measure the treated stand's growth change. A standard method reported in the literature to include this growth change in G&Y predictions consists of using the pairs of observations of yield for treated and non-treated stands and modeling the difference as a function of the time. Typically, the effect of the site quality or Site Index (SI), and the age when the treatment was applied are included as covariates. This resulting function is known as a modifier. Thereby, the modifier can be used to adjust the predicted base-line pattern of growth of a plantation that has not received the silvicultural treatment in a new location. Some researchers have found that the effect of the silvicultural treatment can be modeled using an additive modifier. In this case, the modifier act as an additive term to the base-line yield model. For example, the response of dominant height and basal area growth models for slash pine in the southern U.S. in response to bedding and herbicide treatment could be explained with additive modifiers (Pienaar and Rheney, 1995). The volume responses to woody and herbaceous competition control at age 15 in loblolly pine plantations were also found to be additive (Miller et al., 2003). Multiplicative modifiers have also proven to be useful to adjust the G&Y predictions when thinned (Bailey et al., 1985). Pienaar (1979) proposed that a thinned plantation yield can be obtained applying a multiplicative modifier to the yield estimated from a non-thinning plantation where the age, SI, and trees per acre (TPA) are the same as those remaining from the thinned plantation immediately after thinning. Today, this modifier is also known as the Index of Suppression. However, there are not any recent studies that have validated this procedure for loblolly pine. Other approaches include explicitly incorporating silvicultural treatments into the G&Y as indicator variables (Fang et al., 2001). This approach has two limitations. First, it is not easy to separate the effects of different treatments at the stand level because each silvicultural treatment produces an enhanced growth response of different magnitude and duration. Second, the positive responses in growth due to the implementation of a silvicultural treatment are attributed to improved treatment and a secondary effect due to

stand stocking and structure changes resulting from improved treatment (Miller and Tarrant, 1983). For these reasons, the modifier function is preferred over the use of indicator variables when modeling silvicultural treatment effects. Additionally, modifiers allow flexible response forms corresponding with the varied time-duration effect and intensity of treatments. In this dissertation, a new use of modifiers is proposed. Instead of adopting additive or multiplicative function to adjust the stand yield, the modifiers were applied to the G&Y system's parameters. In this way, the underline model interpretation is unchanged but extensible to include treatment effects.

1.3 Problem Description

Since the silvicultural practices of vegetation competition control, fertilization, and thinning are known to impact the future growth of treated loblolly pine stands, it is crucial to understand how to incorporate these responses and their interactions into growth models. The current loblolly pine models developed by the PMRC and others in the literature tend to use modifiers on the untreated stand-level yield models to predict individual silvicultural treatment conditions. The first limitation with this approach to predict the yield in a new location is the selections of one adequate base-line model. A second limitation is that the temporal treatment effect assumption applied to the yield is not consistent over the long run because the treated stand results in identical conditions of the base-line model. Additionally, so far, all the model-ensembling strategies reported are based on a statistical model selection with few consistent biological support. There is only scarce discussion in the literature on mid-rotation treatment modeling approaches or the potential interaction effect when treatments come together. Even less attention has been made on modeling strategies that use modifiers on instantaneous growth. This research aims to contribute with a better understanding of the mid-rotation treatments modeling and their stand-level effects on the plantation growth over the rotation.

1.4 Objectives and Hypotheses

1.4.1 General Objective

To propose a new stand-level growth and yield system model (taper, basal area, and stand table projection) for loblolly pine plantations in the southern U.S. that addresses the effects of combining mid-rotation silvicultural treatments using the MRT study experiment layout.

1.4.2 Specific Objectives

Objective 1. To propose a new modeling approach for taper that allows for accurate diameter, volume estimation, and uncertainty assessment.

Objective 2. To propose a system of equations that permit prediction and projection of stand-level attributes when mid-rotation silvicultural practices are present using the MRT study experiment layout.

Objective 3. Evaluate the impact on product class distribution when stand management includes mid-rotation silvicultural practices in the MRT study context.

1.4.3 Hypotheses

H1. The assumptions with the actual parametric taper models used so far for loblolly pine results in restricted flexibility, increasing the diameter bias prediction and error prediction compared with the alternative semi-parametric approach.

H2 Fertilization and vegetation competition control used as mid-rotation silvicultural treatments produce a positive response in growth for stand-level attributes like dominant height and basal area.

H3. The interaction of fertilization and vegetation competition control on the stand-level basal area is synergistic.

H4. Traditionally the silvicultural treatment effects are included in G&Y models as modifier equations coupled with yield baseline models. More meaningful models can be obtained if the modifiers are coupled with some of the growth models' parameters.

H5. The proportion of saw timber products (Psw) at rotation age increases when thinning is combined with other silvicultural practice. In general Psw with thinning+CVC+Fertilization > Psw with thinning + Fertilization > Psw with thinning +CVC > Psw with thinning only > Psw without thinning.

1.5 Dissertation Structure

This dissertation compiles the research results in four main chapters. The chapters are writing following general guidelines for preparing scientific manuscripts. Chapter 2 presents a literature review about the most common mid-rotation silvicultural treatments applied in loblolly pine but with an emphasis on modeling strategies. A conceptual framework about the effect of treatments on growth is discussed. Chapter 3 presents a new taper equation for loblolly pine using penalize spline regression. Although semiparametric methods have been used for taper modeling previously, up to my knowledge, this is the first time this method is adapted for loblolly pine. Comparisons with traditionally parametric taper equations are presented and discussed. Taper models are always required when a G&Y system is developed. The tree-wood volume in the MRT sampled plots was estimated with this new taper equation. Chapter 4 proposes a new stand-level dynamic growth system for loblolly pine plantations that can explain stand attributes changes due to silvicultural practices at mid-rotation. The system includes models for dominant height, stand density, and basal area. The driving model for this system is a dominant height's von Bertalanffy differential equation. The

stand density dynamic for thinned and non-thinned conditions is expressed in terms of the dominant height growth. The basal area model is driven by the growth of dominant height and stand density, including explicit parameters-modifiers for each silvicultural practice. Chapter 5 implements a new diameter distribution projection method. The copula theory is explained and linked with the principle of diameter distribution parameter recovery. A new theory for modeling survival diameter distribution is presented. This chapter integrates the taper equation developed in Chapter 3 and the G&Y system presented in Chapter 4.

1.6 Contributions

Most of the material of this dissertation has been previously reported or submitted in the following works. The unpublished material is explicitly mentioned in each chapter.

- Zapata, M., B.P. Bullock, and C.R. Montes (2020). Approaches to model mid-rotation treatment effects in loblolly pine plantations in the southeastern United States, p: 251-257. In: Bragg, D.C., N.E. Koerth, H.A. Gordon, eds. (2020). Proceedings of the 20th Biennial Southern Silvicultural Research Conference. e-Gen. Tech. Rep. SRS-253. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 338 p. Academic publication of public domain.

  The contributions of this manuscript are included in Chapter 2.

- Zapata-Cuartas, M., B.P. Bullock, and C.R. Montes (2020). A taper equation for loblolly pine using penalized spline regression. Manuscript submitted to *Forest Science*.

  The contribution of this paper are included in Chapter 3.

- Zapata-Cuartas, M., B.P. Bullock, C.R. Montes, and M. Kane (in preparation). Dynamic stand growth model system for loblolly pine responding to mid-rotation treatments. To be submitted to *Forests*.

The contribution of this paper are included in Chapter 4.

References

Bailey, R., Borders, B., Ware, K., and Jones, E. (1985). A compatible model relating slash pine plantation survival to density, age, site index, and type and Intensity of thinning. *Forest Science*, 31(1):180–189.

Fang, Z., Bailey, R. L., and Shiver, B. D. (2001). A multivariate simultaneous prediction system for stand growth and yield with fixed and random effects. *Forest Science*, 47(4):550–562.

Fox, T. R., Eric, J. J., and Allen, L. H. (2004). The evolution of pine plantation silviculture in the southern United States. In *Southern forest science: past, present, and future*, volume 075 SRS-75. of *Gen.Tech.Rep*, page 394. U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC.

Fox, T. R., Lee Allen, H., Albaugh, T. J., Rubilar, R., and Carlson, C. A. (2007). Tree nutrition and forest fertilization of pine plantations in the Southern United States. *Southern Journal of Applied Forestry*, 31(1):5–11.

Huggett, R., Wear, D. N., Li, R., Coulston, J., and Liu, S. (2013). Forecasts of forest conditions. In *The Southern Forest Futures Project: technical report*, Forest Service Southern Research Station: General Technical Report SRS-GTR-178, pages 73–101. Wear, D.N., and J.G. Greis, Asheville, NC.

Max, T. and Burkhart, H. (1976). Segmented polynomial regression applied to taper equations. *Forest Science.*, 22(3):283–289.

Miller, J. H., Zutter, B. R., Zedaker, S. M., Edwards, M. B., and Newbold, R. A. (2003). Growth and yield relative to competition for loblolly pine plantations to midrota-

tion—A southeastern United States regional study. *Southern Journal of Applied Forestry*, 27(4):237–252.

Miller, R. and Tarrant, R. (1983). Long-term growth responseof Douglas-fir to ammonium nitrate fertilizer. *Forest Science*, 29(1):127–137.

Pienaar, L. (1979). An approximation of basal area growth after thinning based on growth in unthinned plantations. *Forest Science*, 25(2):223–232.

Pienaar, L. and Rheney, J. (1995). Modeling stand growth and yield response to silvicultural treatments. Technical Report 1995-2, Daniel B. Warnell School of Forest Resources, Universtiy of Georgia, Athens GA.

Prestemon, J. P. and Abt, R. C. (2002). Timber Products Supply and Demand. In *The southern forest resource assessment*, Genera Technical Report GTR-SRS-53, pages 299–325. Wear, D.N., Greis, J.G., Asheville, NC: USDA-Forest Service, Southern Research Station.

Sharma, M., Smith, M., Burkhart, H. E., and Amateis, R. L. (2006). Modeling the impact of thinning on height development of dominant and codominant loblolly pine trees. *Annals of Forest Science*, 63(4):349–354.

Wear, D. N. and Greis, J. G. (2012). The southern forest futures project: summary report. Technical Report SRS-168, Forest Service Southern Research Station, Asheville, NC.

Weiskittel, A. R., editor (2011). *Forest growth and yield modeling*. Wiley, Hoboken, NJ.

Chapter 2

Approaches to model mid-rotation treatment effects in loblolly pine plantations in the southeastern United States[1]

[1]Material presented in academic conference and published in public domain proceedings: Zapata, M., B.P. Bullock, and C.R. Montes (2020). Approaches to model mid-rotation treatment effects in loblolly pine plantations in the southeastern United States, p: 251-257. In: Bragg, D.C., N.E. Koerth, H.A. Gordon, eds. (2020). Proceedings of the 20th Biennial Southern Silvicultural Research Conference. e-Gen. Tech. Rep. SRS-253. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 338 p.

Abstract

Loblolly pine (*Pinus taeda* L.) plantations are extensive in the southeastern United States and represent a significant component of the forest products market in this region. For optimal stand-level management decisions, the growth response to any combination of mid-rotation silvicultural treatments like fertilization, thinning, or competing vegetation management needs to be accurately predicted over the long term. We present a review of the most common mid-rotation silviculture treatments applied in loblolly pine plantations and provide a conceptual framework about their effects. Three theoretical expected types of responses are illustrated as a guide for a better and comprehensive modeling approach selection. Preliminary exploratory data analysis of remeasured plots from a mid-rotation treatment study established by the Plantation Management Research Cooperative at the University of Georgia indicates that basal area growth after thinning follows a Type I response and treatments such as thinning + fertilization or thinning + fertilization + vegetation control can follow either Type I or Type II responses. Finally, some thoughts are provided on how to improve the prediction accuracy of growth and yield models considering mid-rotation treatments.

2.1 Introduction

Throughout the southern U.S., pine plantations have made a significant expansion. Particularly, loblolly pine (*Pinus taeda* L.) plantations represent an important sector of the forest products market in this region. At the same time, improvements in silvicultural practices and intensive management have shown remarkable gain in productivity (Fox et al., 2004). Mid-rotation silvicultural practices have shown important improvements in growth (Snowdon, 2002). The most common silvicultural practices at mid-rotation are thinning, fertilization and vegetation control. However, for optimal stand-level management decisions, the growth response after a specific silvicultural treatment, or combination of treatments, needs to be

accurately predicted over the long-term. Reliable estimates are key for landowners to make informed management decisions or to evaluate investment opportunities.

Existing loblolly pine growth and yield equation systems have been extended with multiplicative or additive sub models, also called "modifiers", at the stand level to represent the changes in growth due to individual silvicultural practices (Pienaar, 1997; Amateis, 2000; Gyawali and Burkhart, 2015). Evidence of interactions between silvicultural treatments has been reported (Sword Sayer et al., 2004), but little research exists that explicitly considers the interaction effect on the growth and yield equations system. Even fewer considerations have been given to the potential type of response that would be observed after a mid-rotation treatment is applied.

Consider the case of modeling the yield of a stand that received two or more mid-rotation silvicultural treatments. The simple addition of independently fitted modifiers (or the sequence of multiplications according to the case) does not guarantee an accurate prediction of the stand yield. A more elaborate procedure is needed in such a way that the resulting interactions of treatments are being considered into growth and yield equation systems. Conventional statistical procedures could be used to estimate the interactions' effect but require long-term field experimental data. To address this issue, the Plantation Management Research Cooperative (PMRC) at the University of Georgia has established and repeatedly measured several controlled loblolly pine mid-rotation silvicultural treatment (MRT) experiments across the southeastern United States. Despite these efforts, not all the potential combinations of type and intensity of silvicultural treatments are available in controlled experiments. Additional information is required to improve the accuracy of the predictions, especially if it will be on sites different from those where the controlled experiments were established.

Alternatively, non-empirical growth models have been proven to perform successfully in explaining the stand level growth responses to silvicultural treatments in loblolly pine

and other species. These non-empirical models are known as process-based models. Simulation outputs of process-based models can be integrated as a priori knowledge with experimental data to improve the precision of growth predictions. Process-based models use well-established physiological models and only require environmental information, which is easily gathered and available for the southeastern United States. Prior research has shown that process-based models can model the productivity effects of mid-rotation treatments in loblolly pine plantations (Bryars et al., 2013; Subedi and Fox, 2016).

This paper reviews the most common mid-rotation silvicultural treatments applied in loblolly pine plantations and presents a conceptual framework about response types as a guide for model development. Also, preliminary results from a PMRC study are used to illustrate the expected nature of the responses in loblolly pine. Finally, some thoughts are provided on how to improve the accuracy in growth and yield predictions when several combinations of mid-rotation silvicultural treatments are present.

2.2 Fertilization

Silvicultural practices serve to increase productivity in loblolly pine plantations. For example, mid-rotation fertilization with nitrogen (N; 150 to 200 pounds per acre) and phosphorous (P; 25 to 50 pounds per acre) on a majority of soils types throughout the southeastern United States produces a large and consistent growth response in the periodic volume growth of approximately 25 percent (Fox et al., 2007). N and P fertilization with responses of approximately 55 cubic feet per acre per year for an 8-year period have been shown to enhance the profitability of loblolly pine plantations (Fox et al., 2007). Fertilization treatments produce an increases nutrient availability at the tree-level, and as a consequence, an increased leaf area index is expected, resulting in improved growth (Albaugh et al., 2017).

2.3 Competition Control

At early-ages herbaceous and hardwood competition can affect adversely gains in growth (Jokela et al., 2010). Early control of competing woody and herbaceous vegetation increases wood volume by 23-121 percent when evaluated at an age of 15 years (Miller et al., 2003). At the tree-level, competition reduces nutrient availability and increase the competition for water, which influences radiation use efficiency.

2.4 Thinning

Thinning is another important silvicultural practice in loblolly pine plantations. There is evidence that thinning not only changes the structure of the stand (diameter distributions, and changes in spatial competition) but also alters the post-thinning height growth of dominant and codominant trees (Sharma et al., 2006) and improves the stem quality of the residual stand. Thinning, in particular, has received considerable attention from researchers. Some studies of thinning effects have evaluated empirical approaches at the tree-level (Soderbergh and Ledermann, 2003; Albaugh et al., 2017) or stand-level (Franklin et al., 2009). Thinning is a way to modify the process of site resource allocation. In general, the primary objective of this treatment is to concentrate light, water, and nutrients on fewer and better trees that can survive over the entire rotation, ending up in a higher valued product.

2.5 Modeling Mid-Rotation Treatments

The extra growth gained at the stand-level (or plot-level) is always referenced or measured with respect to the growth in the untreated stand (called the control plot or untreated plot) with the same site and stand structural attributes, including site index (SI), trees per acre (TPA), and basal area (BA). The treatments can be a single silvicultural practice or

a combination of practices, and the expected response, which may also depend on several other stand characteristics such as age, may have a longer-term effect or a more temporary effect.

Long-term experimentation is the correct approach to infer about the growth drivers and is the only way to evaluate the magnitude of the response given specific treatments. However, these kinds of experiments are costly and time-consuming to install and measure, and it is impossible to observe the response in all possible combinations of experimental factors across the southern United States. Researchers conventionally have used experimental data to update growth and yield equations in order to be able to infer the expected response in a new location. This could involve either explicitly modeling the treatment effect into the growth and yield equations or modeling a response sub-model independently and using it as a modifier of the untreated growth and yield equations.

The explicit modeling of treatments has been done by incorporating silvicultural treatments into the growth and yield models as indicator variables (Fang et al., 2001). This approach has two limitations: (1) it is difficult to separate the effects of different treatments at the stand-level because each silvicultural treatment produces an enhanced growth response of different magnitude and duration, and (2) the positive responses in growth due to the implementation of a silvicultural treatment is attributed both to the improved treatment and a secondary effect of changes to stand stocking and/or structure that result from the improved treatment (Miller and Tarrant, 1983).

The alternative approach consists of estimating a modifier equation. This method is used to adjust the typical or base pattern of growth of a plantation that has not received a given silvicultural treatment (also called the control case). Some researchers have found that the effect of the silvicultural treatment can be modeled using an additive modifier (Amateis, 2000). In this case, the modifier is added to the base growth model. For example, the response of dominant height and basal area growth models for slash pine (*Pinus elliottii*

Engelm.) in the southern United States in response to bedding and herbicide treatment could be explained with additive modifiers (Pienaar and Rheney, 1995). The volume responses to woody and herbaceous competition control at age 15 years in loblolly pine plantations were also found to be additive (Miller et al., 2003).

Multiplicative modifiers have proven useful for growth and yield equations that consider thinning practices (Bailey et al., 1985). For instance, a modifier can be incorporated into a procedure to evaluate the growth of a thinned plantation compared to a non-thinned plantation of the same age, SI, and TPA remaining in the thinned plantation immediately after thinning (Pienaar, 1997). Today, this procedure, also known as the Index of Suppression, continues to be used in growth and yield models with thinning management practices. However, there are no recent studies that have validated this procedure for loblolly pine under more intensive management regimes.

Finally, updated growth and yield models are essential to project the long-term stand-level effects of silvicultural treatments, which help to support forest management decisions. This becomes even more important when one considers that field trials are challenging to replicate and cannot be installed due to their expensive and time-consuming nature (Weiskittel, 2011).

2.6 Response Type Models

A prior understanding of the potential types of stand growth response is important to guide the modeling strategy by defining appropriate modifier functional forms. Two response types to silvicultural treatments applied at the establishment or during early ages have been reported as Type I and Type II (Snowdon, 2002). Here we define three possible types of basal area responses when silvicultural treatments are applied at mid-rotation (Figure 2.1). Type I responses are characterized by a temporary increase in growth rate after treatment compared with the untreated scenario or control, and an extra peak in growth is obtained.

After that, the treated stand growth gradually approaches back to the control. The yields of the treated and untreated stands follow a parallel trajectory (Figure 2.1A). This type of response is likely observed with practices that do not permanently change the site conditions, for example, additions of N-fertilizer.

Response Type II typically occurs when the treatment has a substantial and sustained effect on site properties. The resulting growth over time shows a parallel trajectory with the control growth after treatment. Consequently, a divergent trend in yield is observed when compared with the untreated stand condition (Figure 2.1B). Mid-rotation silviculture practices that produce long term changes in productivity are associated with Type II responses, such as an improvement in nutrient-supply with phosphorous on soils with P deficiencies, irrigation, repeated N-fertilization, or hardwood competition release.

Response Type III is characterized by a temporary increase in growth rate after treatment but is not sustained over time. The initial gain in growth is followed by a decline in growth, which ultimately yields the same as the untreated scenario (Figure 2.1C). This type of response is likely observed when the treatments are applied on fully stocked sites and the intra-specific competition is increased due to the treatment.

Response Types I, II, and III could be combined with a thinning modifier. That is, if the silvicultural treatment is applied at the moment of thinning, then the theoretical response type may be observed on the growth curve for the remaining trees after thinning.

2.7 Case Study

The Mid-Rotation Treatment study has 62 installations throughout the southern U.S. established in non-thinned and thinned plantations. Five 0.75-acre treatment plots were established in each installation and trees in the 0.5-acre interior area were remeasured every two years since treatment was applied. One plot did not receive any treatment of thinning and four plots were assigned the following treatments: thinning only, thinning + fertiliza-

tion (200N + 25P), thinning + vegetation control, and thinning + fertilization + vegetation control.

Let us only consider the measurements from the non-thinned stands. For the sake of the example, we calculate the response as the difference in basal area between the treated plot and the control at each measurement age. We use the untreated plot as a control to compute the responses of the thinned plots. Therefore, we expect to see negative values for the response (Figure 2.2). A negative magnitude in response means that the treatment reduces the stand characteristic that we are analyzing. This happens with the stand-level basal area in the thinning treatment. Interestingly in the three physiographic regions, the observed pattern over time results in a constant response, which implies that the thinned stands are following a Type I response.

The control plot for Thinning + Fertilization and Thinning + Fertilization + Vegetation Control was the Thinning only plot. We observe that some plots do not respond to combinations of treatments (Figure 2.3). Stands that reached their productive potential have been shown to not respond to treatments (Zhao et al., 2016). Plots with an increasing response over time indicate a Type II response. Here we only present the absolute response observed in the measured plots, but statistical modeling will be later developed, and the interactions between treatments will be tested with additional remeasurements.

Defining the expected response type is important for determining the modifier equation form. Depending on the site and stand characteristics, the interaction between treatments likely produces a change in the response type. The next step in our research consists of testing a flexible modifier allowing it accounts for the effects of physiographic region, thinning intensity, amount of vegetation competition at the thinning age. The modifier should also be able to represent either a Type I or Type II growth response.

2.8 Process Based Models

Some authors argue that process-based models per se are still limited concerning management and decision support because of their data and parameterization requirements (Mason and Dzierzon, 2006; Franklin et al., 2009). However, the process-based model 3-PG (Physiological Principles to Predict Growth) developed by Landsberg and Waring (1997) has proven to work very well in predicting long-term thinning responses in Scots pine (*Pinus sylvestris* L.) in Finland (Landsberg et al., 2005); BGC-models (biogeochemical-mechanistic models) have been used with Norway spruce (*Picea abies* (L.) H.K.), and European beech (*Fagus sylvatica*) in Central Europe (Petritsch et al., 2007), and TRIPLEX (a hybrid monthly time-step model of forest growth, carbon, and nitrogen dynamic) has been used for Jack pine (*Pinus banksiana* Lamb.) in Canada (Wang et al., 2011).

The process-based model 3-PG (Landsberg and Waring, 1997) is one of the simplest models readily available. The 3-PG model requires monthly climatic information and site soil characteristics to simulate stand growth. Outputs from this model include stem, foliage, and root biomass, stand volume, leaf area index, basal area, and mean diameters at each monthly step. The structural parameters needed to obtain predictions of biomass and volume in untreated stands throughout the southeastern U.S. has been studied by Bryars et al. (2013).

While 3-PG can be used to simulate thinning treatments (Landsberg et al., 2005), it is not as versatile, in general, as the conventional growth and yield model equation systems. However, 3-PG could provide insight about the expected effect when several silvicultural treatments are combined. That is, once a calibrated 3-PG model is obtained for a particular site, it could be used to study the relative growth change under different silvicultural treatment intensities and timings. A reliable process-based model is valuable, especially to generate information about the expected effect of silvicultural treatments on sites where there

are no existing trials, or to inform empirical models with prior information on interaction effects when few trial replications are available.

2.9 Conclusions

Since the silvicultural practices of vegetation competition control, fertilization, and thinning are known to impact the future growth and yield of loblolly pine stands, it is important to understand how to incorporate these responses and their interactions into growth and yield predictions models. The current loblolly pine models developed by the PMRC, as well as others in the literature, always consider the stand-level growth and yield models for the untreated conditions and use modifiers to adjust the projected growth as a response to individual silvicultural treatments. We illustrate hypothetical possible response types when mid-rotation treatments are present. Preliminary results from a PMRC MRT study show that the growth response of loblolly pine plots with mid-rotation treatments like Thinning + Fertilization or Thinning +Fertilization + Vegetation Control can be either Type I or Type II. Plots with Thinning only follow a Type I growth pattern exclusively. Process-based knowledge promises to be a useful information source that can be incorporated as a piece of prior information in empirical models. Further research will evaluate the prediction accuracy of models combining experimental data and process-based information.

ACKNOWLEDGMENTS Authors would like to thank the members of the Plantation Management Research Cooperative (PMRC) for their support of this research.

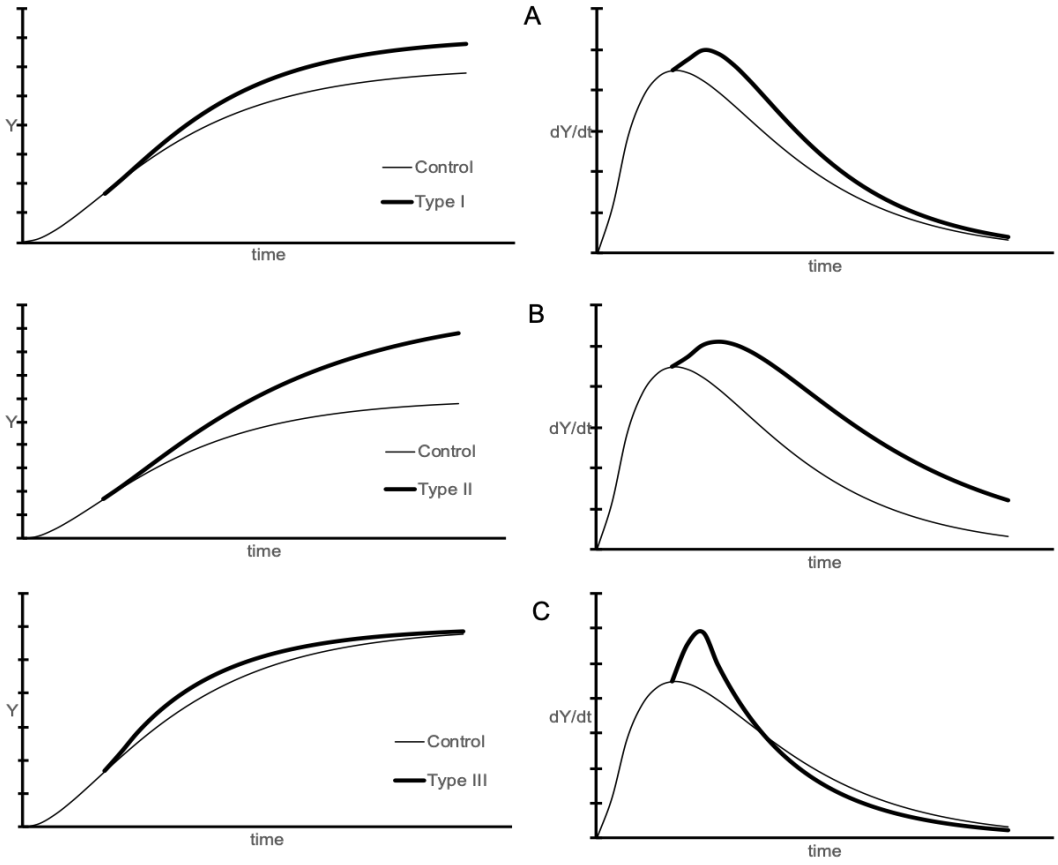

Figure 2.1: A theoretical pattern of responses to mid-rotation silvicultural treatments in yield and growth for the basal area of a loblolly pine plantation. (A) Type I response, (B) Type II response, and (C) Type III response.

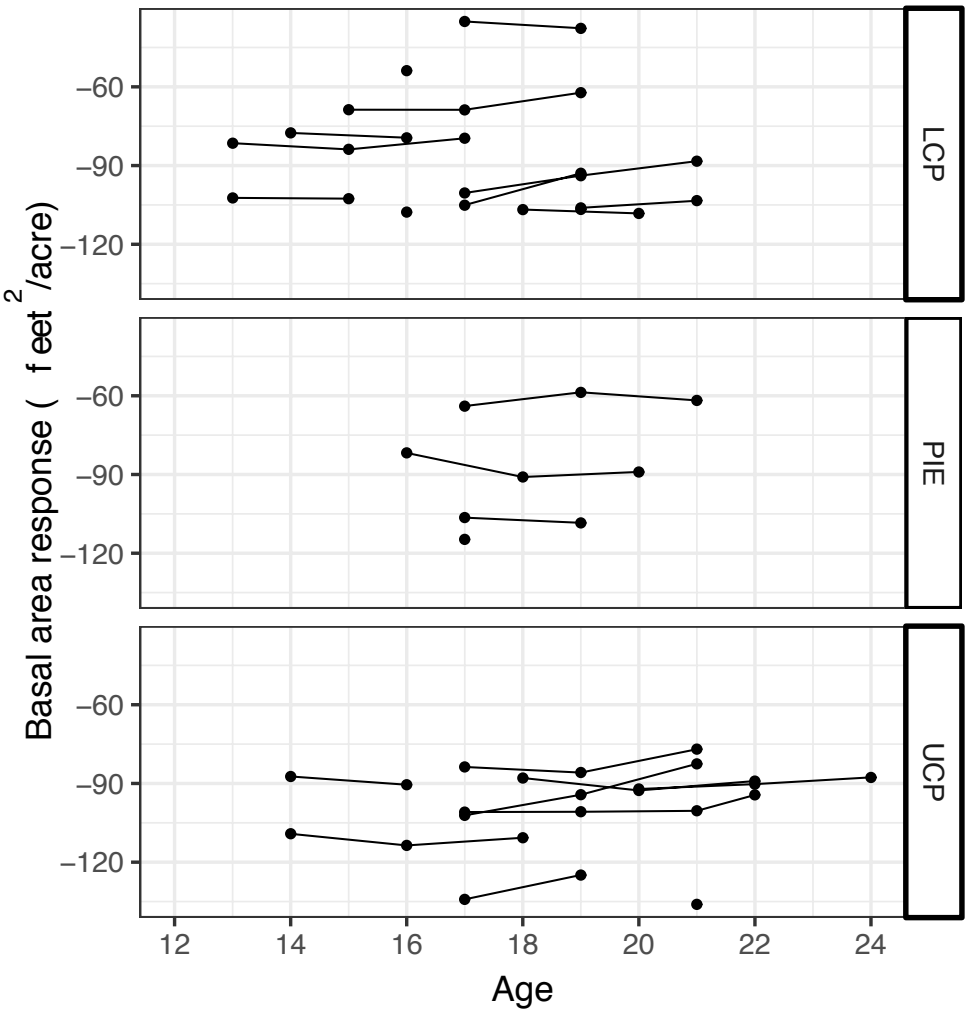

Figure 2.2: Basal area response after thinning for 25 plots of the PMRC MRT study grouped by physiographic region. LCP: Lower Coastal Plain, PIE: Piedmont, and UCP: Upper Coastal Plain. The response at each measurement was defined as the difference in basal area between the thinned plot and the unthinned control plot. The remeasurements of each plot are represented with dots connected by lines.

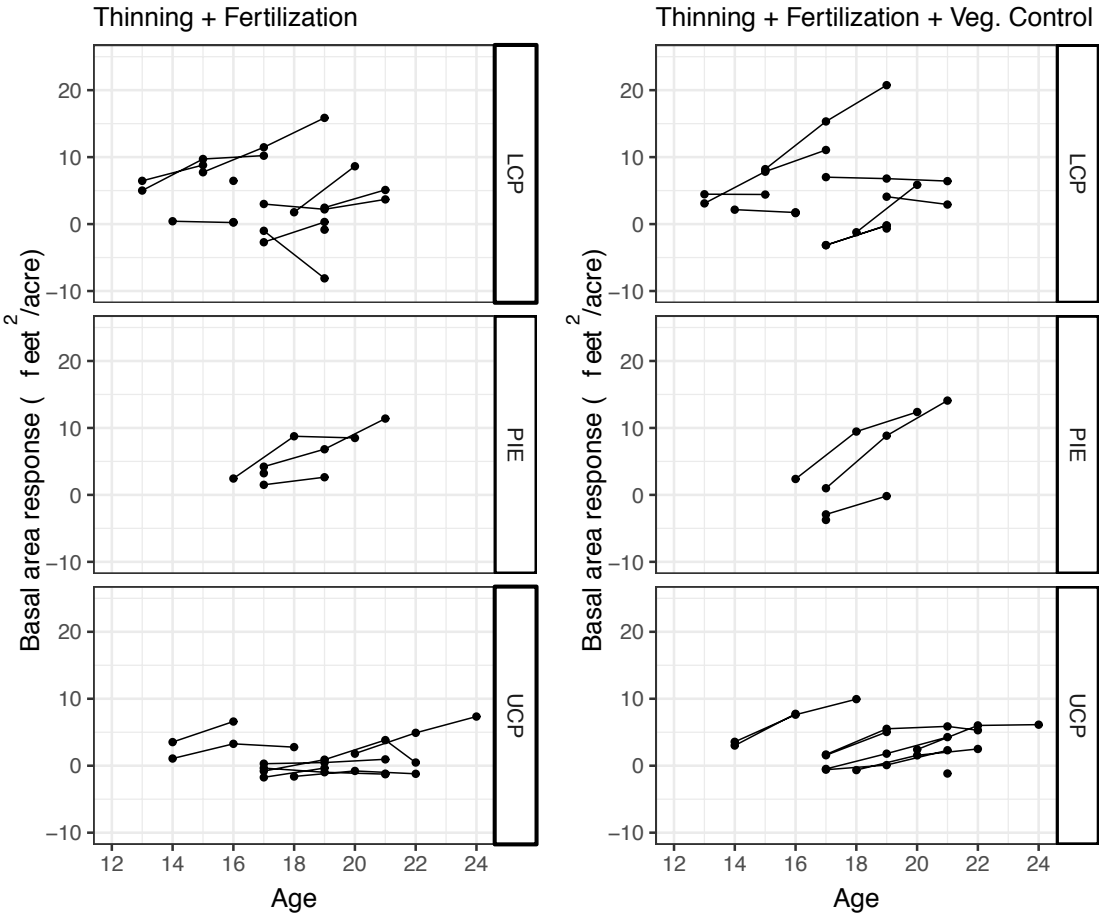

Figure 2.3: Basal area response after thinning and with combinations of treatments: thinning plus fertilization (left) and thinning plus fertilization plus vegetation control (right). LCP: Lower Coastal Plain, PIE: Piedmont, and UCP: Upper Coastal Plain. The response at each measurement was defined as the difference in basal area between the plots with combined treatments and the respective thinning only control plot.

References

Albaugh, T. J., Fox, T. R., Rubilar, R. A., Cook, R. L., Amateis, R. L., and Burkhart, H. E. (2017). Post-thinning density and fertilization affect Pinus taeda stand and individual tree growth. *Forest Ecology and Management*, 396:207–216.

Amateis, R. L. (2000). Modeling response to thinning in loblolly pine plantations. *Southern Journal of Applied Forestry*, 24(1):17–22.

Bailey, R., Borders, B., Ware, K., and Jones, E. (1985). A compatible model relating slash pine plantation survival to density, age, site index, and type and Intensity of thinning. *Forest Science*, 31(1):180–189.

Bryars, C., Maier, C., Zhao, D., Kane, M., Borders, B., Will, R., and Teskey, R. (2013). Fixed physiological parameters in the 3-PG model produced accurate estimates of loblolly pine growth on sites in different geographic regions. *Forest Ecology and Management*, 289:501–514.

Fang, Z., Bailey, R. L., and Shiver, B. D. (2001). A multivariate simultaneous prediction system for stand growth and yield with fixed and random effects. *Forest Science*, 47(4):550–562.

Fox, T. R., Eric, J. J., and Allen, L. H. (2004). The evolution of pine plantation silviculture in the southern United States. In *Southern forest science: past, present, and future*, volume 075 SRS-75. of *Gen.Tech.Rep*, page 394. U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC.

Fox, T. R., Lee Allen, H., Albaugh, T. J., Rubilar, R., and Carlson, C. A. (2007). Tree nutrition and forest fertilization of pine plantations in the Southern United States. *Southern Journal of Applied Forestry*, 31(1):5–11.

Franklin, O., Aoki, K., and Seidl, R. (2009). A generic model of thinning and stand density effects on forest growth, mortality and net increment. *Annals of Forest Science*, 66(8):815–815.

Gyawali, N. and Burkhart, H. E. (2015). General response functions to silvicultural treatments in loblolly pine plantations. *Canadian Journal of Forest Research*, 45:252–265.

Jokela, E. J., Martin, T. A., and Vogel, J. G. (2010). Twenty-Five Years of Intensive Forest Management with Southern Pines: Important Lessons Learned. *Journal of Forestry*, 108:338–347.

Landsberg, J., Makela, A., Sievanen, R., and Kukkola, M. (2005). Analysis of biomass accumulation and stem size distributions over long periods in managed stands of Pinus sylvestris in Finland using the 3-PG model. *Tree Physiology*, 25(7):781–792.

Landsberg, J. and Waring, R. (1997). A generalized model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. *Forest Ecology and Management*, 95(3):209–228.

Mason, E. G. and Dzierzon, H. (2006). Applications of modeling to vegetation management. *Canadian Journal of Forest Research*, 36(10):2505–2514.

Miller, J. H., Zutter, B. R., Zedaker, S. M., Edwards, M. B., and Newbold, R. A. (2003). Growth and yield relative to competition for loblolly pine plantations to midrotation—A southeastern United States regional study. *Southern Journal of Applied Forestry*, 27(4):237–252.

Miller, R. and Tarrant, R. (1983). Long-term growth responseof Douglas-fir to ammonium nitrate fertilizer. *Forest Science*, 29(1):127–137.

Petritsch, R., Hasenauer, H., and Pietsch, S. A. (2007). Incorporating forest growth response to thinning within biome-BGC. *Forest Ecology and Management*, 242(2-3):324–336.

Pienaar, L. and Rheney, J. (1995). Modeling stand growth and yield response to silvicultural treatments. Technical Report 1995-2, Daniel B. Warnell School of Forest Resources, Universtiy of Georgia, Athens GA.

Pienaar, L. V. (1997). An approximation of basal area growth after thinning based on growth in unthinned plantations. *Forest Science*, 25(2):223–232.

Sharma, M., Smith, M., Burkhart, H. E., and Amateis, R. L. (2006). Modeling the impact of thinning on height development of dominant and codominant loblolly pine trees. *Annals of Forest Science*, 63(4):349–354.

Snowdon, P. (2002). Modeling Type 1 and Type 2 growth responses in plantations after application of fertilizer or other silvicultural treatments. *Forest Ecology and Management*, 163:229–244.

Soderbergh, I. and Ledermann, T. (2003). Algorithms for simulating thinning and harvesting in five European individual-tree growth simulators: a review. *Computers and Electronics in Agriculture*, 39(2):115–140.

Subedi, S. and Fox, T. R. (2016). Modeling repeated fertilizer response and one-time midrotation fertilizer response in loblolly pine plantations using FR in the 3-PG process model. *Forest Ecology and Management*, 380:90–99.

Sword Sayer, M., Goelz, J., Chambers, J., Tang, Z., Dean, T., Haywood, J., and Leduc, D. (2004). Long-term trends in loblolly pine productivity and stand characteristics in response to thinning and fertilization in the West Gulf region. *Forest Ecology and Management*, 192(1):71–96.

Wang, W., Peng, C., Zhang, S., Zhou, X., Larocque, G. R., Kneeshaw, D. D., and Lei, X. (2011). Development of TRIPLEX-Management model for simulating the response of forest growth to pre-commercial thinning. *Ecological Modelling*, 222(14):2249–2261.

Weiskittel, A. R., editor (2011). *Forest growth and yield modeling*. Wiley, Hoboken, NJ.

Zhao, D., Kane, M., Teskey, R., Fox, T. R., Albaugh, T. J., Allen, H. L., and Rubilar, R. (2016). Maximum response of loblolly pine plantations to silvicultural management in the southern United States. *Forest Ecology and Management*, 375:105–111.

Chapter 3

A Taper Equation for Loblolly Pine Using Penalized Spline Regression [1]

[1] Zapata-Cuartas, M., B.P. Bullock, and C.R. Montes. A taper equation for loblolly pine using penalized spline regression. Accepted by *Forest Science*. Reprinted here with permission of publisher.

Abstract

Stem profile needs to be modeled with an accurate taper equation to produce reliable tree volume assessments. We propose a semiparametric method where few *a priori* functional form assumptions or parametric specification is required. We compared the diameter and volume predictions of a penalized spline regression (P-Spline), P-Spline extended with an additive dbh-class variable, and six alternative parametric taper equations including single, segmented, and variable-exponent equation forms. We used taper data from 147 loblolly pine (*Pinus taeda* L.) trees to fit the models and make comparisons. Here we show that the extended P-Spline outperforms the parametric taper equations when used to predict outside bark diameter in the lower portion of the stem, up to 40% of the tree height where the more valuable wood products (62% of the total outside bark volume) are located. For volume, both P-Spline models perform equal or better than the best parametric model, Max and Burkhart (1976) with taper calibration, which could result in possible savings on inventory costs by not requiring an additional measurement. Our findings suggest that assuming *a priori* fixed form in taper models imposes restrictions that fail to explain the tree form adequately compared with the proposed P-Spline.

Management and Policy Implications

Our semiparametric fitting approach translates into more precise stem diameter predictions compared with traditional parametric taper equations. In terms of volume, we show that the proposed method is flexible enough to accurately predict total and merchantable volume similar to calibrated segmented equations. Accurate diameter predictions and volume estimations at the tree-level can add value to the data inventory process and reporting by reducing bias and allowing better management decisions for the most planted species in the southern United States. Although our sample size is only 147 trees, the method described

here produces excellent volume prediction qualities without requiring calibration with extra upper-stem diameter measurement.

3.1 Introduction

Foresters have extensively used taper modeling in plantation management as it enables total and merchantable volume estimations at the tree and stand-level using information from standard forest inventories (Kozak et al., 1969). Typically, tree taper is expressed as a mathematical stem-profile equation that, when integrated, derives an expression of total or merchantable volume depending on the specified upper diameter limit. Tree-level volume estimation is necessary for timber resource evaluation, to evaluate the effect of silvicultural treatments (Haywood, 2005), and to evaluate genetic gains in improved seedlings (Wood et al., 2015). Therefore, gains on accurate stem taper predictions will result in more exact volume estimations and allow for better management decisions to be made.

The typical predictors used in taper equations are diameter at breast height (dbh) and total stem height (H). Different predictors combined with dbh and H have been evaluated in the past to try to increase volume prediction accuracy, e.g., crown length (Leites and Robinson, 2004), or an additional one or two upper stem diameters at fixed locations on the stem (Trincado and Burkhart, 2006; Cao and Wang, 2011). Nevertheless, if there are measurement errors higher than 5% in the upper stem variable, the estimates of the volume will be biased and may not be worth the additional upper stem measurement effort (Arias-Rodil et al., 2017).

Different equation forms have been proposed to model taper, ranging from simple linear models to highly non-linear systems of equations. Regardless of its complexity, any taper equation can be classified in one of the following groups (Diéguez-Aranda et al., 2006): single equations, segmented taper models, and variable-exponent taper models. Single equations describe the entire profile of the stem using a unique equation. Segmented taper models

use different equations with join points or knots to guarantee continuity and assume that each model explains a geometric shape present along the stem. Variable-exponent taper models are characterized by an exponent expression that changes along the stem to represent different shapes like neloid, paraboloid, or conoids.

Although varying taper equation forms and prediction combinations exist, most of the empirical taper equations published simplify the relationship between the diameter at given tree height and the predictors to be parametric, meaning that *a priori* fixed functional form, based on an ideal shape and a finite set of parameters, is assumed to describe the tree taper adequately.

Such parametric representations can be restrictive, and rarely is a unique functional form correct; instead, it can lead to severely biased estimates and misleading inferences in taper modeling. Alternatively, penalized spline regression (P-Spline) combines parametric and nonparametric principles conveniently. Even though the use of a relatively large number of parameters is less parsimonious, it allows for flexible fitting, controlling smoothness using the mixed-effect model representation (Ruppert et al., 2003, 2009).

Tree taper studies using simple splines or P-Splines are not new. Smoothing splines were initially used to obtain accurate taper curves (Goulding, 1979; Liu, 1980). Figueiredo-Filho et al. (1996) and Kuzelka and Marusák (2015) studied the effect of the amount and distribution of measured diameters on the accuracy of stem curves using different types of splines. Koskela et al. (2006) used cubic smoothing splines as an interpolation method from observed and estimated stem diameters to determine optimal cutting points along the stem in harvesting situations. Kuzelka and Marusák (2014) compared two types of spline regression (smoothing splines and B-Spline) with four parametric taper equations. They found no differences in accuracy between a one variable-exponent model, and the spline models.

Lappi (2006) used smoothing splines to estimate means, variances, and correlations of stem dimensions, which are defined at specific positions along the stem and relative to the

diameter at breast height. Lappi's model includes four diameter measurements at absolute heights below the dbh and eight diameters at relative distances from the breast height to the treetop. Finally, the taper is recovered interpolating between stem dimensions. Kublin et al. (2008) improved upon Lappi's model proposing a functional regression model for diameter prediction on a continuous measurement grid instead of using a set of discrete measurements as was recommended by Lappi. They assume that observed diameters come from independent realizations of a smooth random function with unknown mean and covariance functions. However, they are estimable with a sample of trees measured at irregular grids. They also modeled the within-tree empirical covariance for fixed dbh-classes and fixed within-tree diameter positions (10, 30, 50, and 90%) using multivariate tensor splines. Kublin et al. (2013) presented a semiparametric cubic B-Spline mixed-effect model where the population mean taper curve, as well as the tree-specific random deviation, are a smooth function of relative height. According to Kublin et al. (2013), the prediction model for the jth diameter $D(h_{ij})$ from the ith tree at the relative height h, $j = 1, \ldots, n_i$, $i = 1, \ldots, n$ can be represented as

$$D(h_{ij}) = f(h_{ij}) + g_i(h_{ij}) + \epsilon_{ij} \tag{3.1}$$

where $f(h_{ij})$ is a B-Spline population mean function, $g_i(h_{ij})$ is the random tree specific deviation from $f(h_{ij})$, which is assumed to be a Gaussian process with mean 0 and covariance function depending on the within-tree location, and ϵ_{ij} is a vector of residual errors assumed to be normally distributed. A similar idea was presented by Krivobokova et al. (2008), they used a hierarchical mixed-effect model with a spatial adaptive penalty imposing a functional structure on the variance for the smoothing parameter. Scolforo et al. (2018) used a truncated quadratic spline basis to analyze differences in taper for genetic families of *Eucalyptus* L'Hér planted in Brazil.

In the semiparametric estimation, depending on the constraint imposed on the smoothing parameter, the estimator moves from one situation of misspecification (biased but low noise)

to one with a constraint-free nonparametric (unbiased but highly noised models), which is called the bias-variance trade-off (Gu, 2013). Ruppert et al. (2003) showed that providing an adequate distribution of knots, their number and position does not substantially affect the quality of smoothing as much as the changes on the smoothing parameter does, and most important is that given a set of knots with fixed locations the representation of penalizing spline through Restricted Maximum Likelihood method (REML) automatically chooses an optimal smooth estimator considering the bias-variance trade-off. In this paper, we use the term B-Spline to refer to spline methods defined on recurrence relation (see de Boor (2001)) and P-Spline for those splines that use truncated cubic basis functions.

Previous works comparing the performance of parametric and semiparametric models for taper modeling showed that semiparametric-based taper equations could be highly efficient for describing both stem form and volume. Scolforo et al. (2018) found that a quadratic penalized regression with an additional size-independent variable produces accurate and stable predictions along the tree when comparing with alternative polynomial, segmented, and variable exponent taper equation in four genetic families of eucalyptus in Brazil. Robinson et al. (2011) found similar accuracy behavior when compare the flexible variable exponential equations "Model 02" proposed by Kozak (2004) and a generalized additive model (Hastie and Tibshirani, 1986) with cubic spline as smooth functions for six British Columbia species groups.

This study aimed to evaluate the performance of semiparametric-based P-Spline methods to provide accurate estimations of stem diameter and volume for loblolly pine. We compare six contrasting well accepted parametric taper equations with a new proposed P-Spline model. The specific research objectives were to (1) describe the use of P-Spline methods for modeling the stem taper curve, and (2) compare the accuracy in diameter and volume prediction by relative height sections along the stem and for the total tree.

We expect that P-Splines would prove to be attractive for precise diameter and merchantable volume predictions compared with the traditional parametric approach because this technique is closely related to smoothing. We are particularly interested in loblolly pine because its plantations are extensive in the southern United States and represent a significant component of the forest products market in this region (Huggett et al., 2013). A far as we know, P-Splines have not yet been evaluated as alternative taper modeling in loblolly pine.

3.2 Materials and Methods

3.2.1 Study Area and Data Collection

We used taper data from 147 loblolly pine trees extracted from non-thinned planted stands in the Whitehall Forest at the University of Georgia (central geographic coordinates for the sample sites: 33.8842, -83.3552) during the fall seasons from 2014 to 2018. We selected the sample from stands with different ages, site indices, and management regimens, making sure to avoid trees with severe anomalies in their profile, ensuring a wide range in dbh and H. The sample trees were felled and bucked into approximately 1.2 m sections starting from the ground and including measurements at 0.15 m, 0.6 m, and at breast height at 1.37 m. Diameters outside bark (*dob*) were recorded at each measured height section (h) to the nearest 0.03 cm. The uppermost section of the stem was determined by a top *dob* of 7.62 cm, and the distance from the top to the tip was measured. The dbh and H were also recorded for each tree before being felled. The volume of each section was calculated using Smalian's formula (Burkhart et al., 2019, p. 70), except for the top section, which was assumed to have the volume of a cone. Finally, the observed total tree volume (V) was calculated as the sum of all its sections.

A summary of statistics for dbh, H, and V of sample trees are reported in Table 3.1.

3.2.2 Penalized Spline Regression

We propose a penalized spline linear regression, where we assume that the relative mean taper curve is a smooth spline function of degree $p = 3$. The spline function consists of a polynomial equation and a set of positive truncated basis, which can be represented as a standard linear mixed-effect (LME) model

$$\begin{aligned} dr_{ij} &= \beta_0 + \beta_1 q_{ij} + \beta_2 q_{ij}^2 + \beta_3 q_{ij}^3 + \sum_{k=1}^{K} \beta_{3+k,ij} [(q_{ij} - \kappa_k)]_+^3 + \epsilon_{ij} \\ \epsilon_{ij} &= (\epsilon_{i1}, \ldots, \epsilon_{in_i})^T \sim N(0, R) \end{aligned} \tag{3.2}$$

where dr_{ij} is the relative stem diameter in the ith tree ($i = 1, \ldots, 147$) at the jth relative height ($j = 1, \ldots, n_i$), the relative diameter was defined as dob_{ij}/dbh_i, relative height was defined as $q_{ij} = h_{ij}/H_i$, H_i is the total height of the ith tree, h_{ij} is the jth measured height along the stem for the ith tree, β_0 , β_1, β_2, and β_3 are fixed parameters associated with the polynomial part of the model. $\kappa_1, \ldots, \kappa_K$ is a sequence of K given knots located in the relative height interval $[0, 1]$, which are linked points for each truncated cubic spline basis. The term $[(q_{ij} - \kappa_k)]_+^3$ introduces a positive cubic spline basis function into the model such that for a given q_{ij}, the basis takes a value of zero when $q_{ij} < \kappa_k$ and $(q_{ij} - \kappa_k)^3$ in other cases. $\beta_{3+k,ij}$ are considered random coefficients that account for deviation from the population mean taper curve. The vector of residual errors ϵ_{ij} is assumed to follow a normal distribution with covariance matrix R. Eq. 3.2 returns a smooth mean stem profile in relative scales and defines a simple P-Spline model. However, extra variability could remain because of the range of tree sizes or ages. Our data set does not cover a wide range of ages. Therefore, we proposed to expand Eq. 3.2 with an additive dbh-class variable. The second P-Spline proposed model is

$$dr_{ijl} = \beta_0 + \alpha_l + \beta_1 q_{ij} + \beta_2 q_{ij}^2 + \beta_3 q_{ij}^3 + \sum_{k=1}^{K} \beta_{3+k,ij}[(q_{ij} - \kappa_k)]_+^3 + \epsilon_{ijl} \quad (3.3)$$

where α_l is the additive parameter effect for the lth dbh-class ($l = 1, \ldots, m$). We used dbh-class as a proxy for tree-size, and we hypothesize that an improvement can be achieved by allowing the mean smooth response shifting the prediction according to the tree size. The dbh-class were defined by breaking the dbh tree list in 10 equal-size groups ($m = 10$). Inferences with Eq. 3.3 might reveal that relative taper shapes can vary according to the tree size.

Given a set of knots at known locations, a fixed mean response solution for Eq. 3.3 could be found using constrained ordinary least-squares methods (see Hastie (1996)) such that we only need to find $\boldsymbol{\beta}$ that minimize $||\mathbf{dr} - \mathbf{X}\boldsymbol{\beta}||^2 + \lambda^{2p}\boldsymbol{\beta}^T\mathbf{D}\boldsymbol{\beta}$, where $\lambda \geq 0$ is the smoothing parameter, $\mathbf{D} = diag(0_{p+m}, 1, \ldots, 1)_{(K+p+m)\times(K+p+m)}$, and $\lambda^{2p}\boldsymbol{\beta}^T\mathbf{D}\boldsymbol{\beta}$ is the roughness penalty. Thus, the ordinary least-squares estimate parameters are $\hat{\boldsymbol{\beta}}_\lambda = (\mathbf{X}^T\mathbf{X} + \lambda^{2p}\mathbf{D})^{-1}\mathbf{X}^T\mathbf{y}$. However, an additional cross-validation method is necessary for choosing λ. Instead of using ordinary least-squares, we use REML for fitting. It allows finding the best linear predictor of model parameters and the smoothing parameter (λ) simultaneously in one step, avoiding the extra cross-validations.

For a complete specification of a P-Spline model, choices of the amount and relative position of the knots have to be made. Contrary to the parametric segmented taper models, P-Spline is flexible enough to permit any amount of internal knots while conserving numerical tractability. We propose to localize the knots using the sample distribution of relative heights taken at K percentiles. Therefore, the locations will be adaptive according to the intensity and availability of measurements along the sampled tree stems. We found the minimum necessary number of knots through a cross-validation method with the available sample data.

3.2.3 Estimation

In this section, we explain the estimation method and inference for the P-Spline in Eq. 3.3. A similar method would apply for Eq. 3.2 omitting the additive variable dbh-class. Once a finite set of internal knots $\kappa_k \in [0,1]$ is provided, the model in Eq. 3.3 is the standard LME model representation for our smoothed taper equation. To simplify the notation, let's express it in matrix form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \epsilon, \quad Cov \begin{bmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}, \tag{3.4}$$

where $\mathbf{y}$ is a vector of observed relative diameters, $\boldsymbol{\beta} = [\beta_0, \alpha_2, \ldots, \alpha_m, \beta_1, \beta_2, \beta_3]$, $\mathbf{u} = [\beta_{3+1}, \ldots, \beta_{3+K}]$ is a vector of random coefficients i.i.d. $N(\mathbf{0}, \mathbf{G})$, and

$$\mathbf{X} = \begin{bmatrix} 1 & I_{1,2} & \ldots & I_{1,m} & q_{1,1} & q_{1,1}^2 & q_{1,1}^3 \\ \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & I_{n,2} & \ldots & I_{n,m} & q_{147,n_{147}} & q_{147,n_{147}}^2 & q_{147,n_{147}}^3 \end{bmatrix},$$

$$\mathbf{Z} = \begin{bmatrix} [(q_{1,1} - \kappa_1)]_+^3 & \ldots & [(q_{1,1} - \kappa_K)]_+^3 \\ \vdots & \ddots & \vdots \\ [(q_{147,n_{147}} - \kappa_1)]_+^3 & \ldots & [(q_{147,n_{147}} - \kappa_K)]_+^3 \end{bmatrix}$$

$I_{i,l}$ is an indicator variable that takes the value of 1 when the ith tree belongs to the lth dbh-class. Let $\mathbf{V} = Cov(\mathbf{y}) = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$. This representation allows us to use REML methods to fit the P-Spline model in Eq. 3.3. Let $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{u}}$, $\hat{\mathbf{R}} = \mathbf{R}(\hat{\theta}_R)$ and $\hat{\mathbf{G}} = \mathbf{G}(\hat{\theta}_G)$ (e.g., $\hat{\mathbf{G}} = \hat{\sigma}^2_{\mathbf{u}}\mathbf{I}$) be the estimated parameters and variance-covariance components for residuals and random effects with respective variance-parameters $\hat{\theta}_R$ and $\hat{\theta}_G$, so that $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{y}$, and $\hat{\mathbf{u}} = \hat{\mathbf{G}}\mathbf{Z}^T\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$. The variance-covariance matrix of

the error term can be expressed in a more general form such that it accounts for within-tree correlation as $\mathbf{R}_i(\hat{\theta}_R) = \sigma^2_\epsilon \Omega_i^{1/2} \Gamma_i(\hat{\theta}_R) \Omega_i^{1/2}$ where $\Omega_i^{1/2}$ is a $(n_i \times n_i)$ diagonal matrix that characterizes tree variance and $\Gamma_i(\hat{\theta}_R)$ is a $(n_i \times n_i)$ matrix that describe the correlation pattern within the measurements for the ith tree. Fitted relative diameters values were obtained from $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}}$, and the smoothing parameter was found as $\lambda = \hat{\sigma}^2_\epsilon / \hat{\sigma}^2_{\mathbf{u}}$.

The representation above allows us to fit a penalized spline smoothing using classical parametric estimation. We use the R-package *nlme* (Pinheiro et al., 2018) to fit the model in Eq. 3.4.

Finding the Number of Knots

Define S as a sequence $S = [2, 4, 6, 8, \ldots, 20]$ of possible realizations for K (number of knots for the P-Spline model). The cross-validation leave-one-out procedure was carried out for each value in S to find which value of K produces the lowest average prediction error. We use the root mean square difference ($RMSD$) as proposed by Stage and Crookston (2007). The $RMSD$ for the ith tree with dbh-class l was defined as

$$RMSD_i = \sqrt{\frac{\sum_{j=1}^{n_i} (d_{ij} - \hat{dr}_{(-i)jl} * \text{dbh}_i)^2}{n_i}} \tag{3.5}$$

where d_{ij} is the jth *dob* measurement in the ith tree, and $\hat{dr}_{(-i)jl}$ is the respective predicted relative diameter outside bark using a fitted model without the ith tree. $RMSD$ was calculated for each tree. The overall difference for each value in S was obtained as the average $RMSD$ over all the tree measurements. The minimum number of S that allows the lowest difference was selected as an adequate number of knots K for the proposed smooth taper model.

3.2.4 Inference on the Mean Taper

We are particularly interested in inferences about $\hat{dr}_{ijl}$, which is a point estimator of dr_{ijl}. Inference about $\hat{dr}_{ijl}$ is based on the conditional distribution $y|\mathbf{u}$ (Ruppert et al., 2003). An approximate $100(1-\alpha)\%$ confidence interval for $E\{dr_{ijl}|\mathbf{u}\}$ in some dbh-class and accounting for bias is

$$\hat{dr}_{ijl} \pm t_{(1-\alpha/2;df)} \hat{sd}\{dr_{ijl} - \hat{dr}_{ijl}\} \tag{3.6}$$

where $\hat{sd}\{dr_{ijl} - \hat{dr}_{ijl}\}$ can be obtained from

$$\hat{sd}\{dr_{ijl} - \hat{dr}_{ijl}\} = \hat{\sigma}_\epsilon \sqrt{\mathbf{C}_q(\mathbf{C}^T\mathbf{C} + \frac{\hat{\sigma}^2_\epsilon}{\hat{\sigma}^2_\mathbf{u}}\mathbf{D})^{-1}\mathbf{C}_q^T} \tag{3.7}$$

where $\mathbf{C}_q = [1, \text{dbh-class}_l, q, q^2, q^3, (q-\kappa_1)^3_+, \ldots, (q-\kappa_K)^3_+]$, is the vector of values for a particular prediction of dr_{ijl} at relative height q for a tree with dbh-class l, and $\mathbf{C} = [\mathbf{C}_{q_i}]_{1\leq i\leq n}$ comes from the regression design matrix. A prediction interval for a new observation at q can be obtained using $\hat{sd}_p$ as

$$\hat{sd}_p\{dr_{ijl} - \hat{dr}_{ijl}\} = \hat{\sigma}_\epsilon \sqrt{1 + \mathbf{C}_q(\mathbf{C}^T\mathbf{C} + \frac{\hat{\sigma}^2_\epsilon}{\hat{\sigma}^2_\mathbf{u}}\mathbf{D})^{-1}\mathbf{C}_q^T} \tag{3.8}$$

The prediction interval is wider because it considers the uncertainty due to the variation of a new observation about its mean and uncertainty regarding that mean. Thus, statements about uncertainty for a predicted diameter to a given height could be reported using $100(1-\alpha)\%$ pointwise confidence or prediction intervals based on Eq. 3.7 or Eq. 3.8 respectively.

3.2.5 Volume Estimation

Calculations of stem volume require knowing the cross-sectional area of the taper function at any relative height q_{ij} as $\frac{\pi}{40000}(dr_{ijl}\text{dbh}_i)^2$. Assuming that the dbh and total tree height H are given and measured without errors, the total stem volume can be calculated as

$$V_i = c \int_0^1 [dr_{ijl}]^2 \, dq \tag{3.9}$$

where $c = \pi 2.5e^{-5}(\text{dbh}_i)^2 H_i$, and dr_{ijl} is the mean relative diameter at height q_{ij} ($q_{ij} \in [0,1]$) as in Eq. 3.2 or Eq. 3.3. We know from the variance definition that $E[dr_{ijl}^2] = \{E[dr_{ijl}]\}^2 + var[dr_{ijl}]$, $\hat{dr}_{ijl}$ provide estimates for $E[dr_{ijl}]$, $(\hat{sd}_p\{dr_{ijl} - \hat{dr}_{ijl}\})^2$ provide estimates for $var[dr_{ijl}]$, the volume as defined in Eq. 3.9 is a continuous monotone integrable function, and their expectation produces iterated integrals, then

$$\begin{aligned} E[V_i] &= c \int_0^1 E[dr_{ijl}^2] \, dq \\ &= c \times \left[\int_0^1 \hat{dr}_{ijl}^{\,2} \, dq + \int_0^1 (\hat{sd}_p\{dr_{ijl} - \hat{dr}_{ijl}\})^2 \, dq \right] \end{aligned} \tag{3.10}$$

Note that if the variance for $\hat{dr}_{ijl}$ in Eq. 3.10 is ignored, then biased estimates of total volume are obtained. Given an estimated taper equation, and a particular dbh-class, the integral term in Eq. 3.10 (without considering the constant term c), will be constant when used for volume prediction with any possible combination of dbh and H in the respective dbh-class. Then it resembles a tree form factor (TFF) for the dbh-class. Consequently, the total volume computation can be sped up when applied to large inventory data sets if the integrals are resolved beforehand as TFF. We calculated the merchantable volume using numerical integration from zero to any upper stem height (h_{ij}/H_i).

3.2.6 Inference on Volume [2]

The conditional variance of volume prediction is obtained by a stochastic integral (Gregoire et al., 2000; Lappi, 2006)

$$var[V_i] = var\left[c\int_0^1 [dr_{ijl}]^2\, dq\right] = c\int_0^1\int_0^1 cov\left\{[dr_{1,ijl}]^2, [dr_{2,ijl}]^2\right\} dq_1 dq_2 \tag{3.11}$$

Variance and covariances of $[dr_{ijl}]^2$ are not available directly from our estimated model, but approximated estimates are obtained by assuming that diameters follows a multivariate normal distribution. Kublin et al. (2013) showed that

$$\begin{aligned} var[V_i] = c^2\int_a^b\int_a^b \quad & var\left[dr_{1,ijl}\right] var\left[dr_{2,ijl}\right] + \\ & 2\left(cov[dr_{1,ijl}, dr_{2,ijl}]\right)^2 + \\ & 4E\left[dr_{1,ijl}\right] E\left[dr_{2,ijl}\right] cov\left[dr_{1,ijl}, dr_{2,ijl}\right] dq_1 dq_2 \end{aligned} \tag{3.12}$$

Using the mixed-effect model framework, the variance and covariance matrix for the estimated parameters can be obtained from $VCOV = \left(\mathbf{C}^T\mathbf{C} + \frac{\hat{\sigma}^2_\epsilon}{\hat{\sigma}^2_{\mathbf{u}}}\mathbf{D}\right)^{-1}$. Given the measurements, the estimated covariance for the mean prediction between q_1 and q_2 can be estimated as $cov\left[\hat{dr}_{1,ijl}, \hat{dr}_{2,ijl}\right] = \mathbf{C}_{q_1} VCOV \mathbf{C}^T_{q_2}$. The two-dimensional integral has to be solved numerically, first integrating the inner integral for each q_j, and then integrating the outer integral. A $100(1-\alpha)\%$ confidence interval for volume is

$$V_i \pm z_{\alpha/2}(var[V_i])^{0.5} \tag{3.13}$$

[2] This sub-section was not included in the submitted paper

3.2.7 Alternative Parametric Taper Equations and Validation

Six alternative parametric taper equations were fit to the data from 147 sampled trees. A cross-validation procedure was performed to assess stem diameter outside bark predictions and volume outside bark predictions of each model, and compared with the proposed P-Spline (Eq. 3.2 and Eq. 3.3). We selected two published equations of each group: single equation, segmented equation, and variable-exponent equation. Next, we briefly describe each of these models.

Sharma and Oderwald (2001) derived a single taper equation by applying dimensional analysis to obtain dimensional compatible taper and volume equations. The taper equation is analytically consistent with the volume equation so that the integration of the taper equation over the total height gives the same result as the volume equation. The taper equation is

$$dob_{ij}^2 = \mathrm{dbh}_i^2 \left(\frac{h_{ij}}{h_D}\right)^{2-\beta_1} \left(\frac{H_i - h_{ij}}{H_i - h_D}\right) \tag{3.14}$$

where h_D is the breast height, β_1 is a parameter related to the taper form. The units for *dob*, dbh, h, h_D, and H are in meters. The corresponding volume equation and merchantable volume equation are

$$V_i = \beta_2 \mathrm{dbh}_i^{\beta_1} H_i^{3-\beta_1} \tag{3.15}$$

$$V_{iu} = \beta_2 \mathrm{dbh}_i^{\beta_1} H_u^{3-\beta_1} \left[(4-\beta_1) - (3-\beta_1) H_u H_i^{-1}\right] \tag{3.16}$$

β_1 and β_2 are unknown parameters and need to be estimated simultaneously to achieve numerical consistency between the tree taper and volume equations. V_{iu} is the merchantable volume for the ith tree up to the height H_u. The simultaneous fitting of Eq. 3.14 and Eq. 3.15 was performed with the R-package *systemfit* (Henningsen and Hamann, 2007).

Kozak et al. (1969) derived a single taper equation to describe the stem shape assuming that the relationship between relative diameter with height follows a parabolic function and imposed constraints to ensure that the estimated *dob* be exactly zero when h equals H. The equation is

$$\left(\frac{dob_{ij}}{\text{dbh}_i}\right)^2 = \beta_1\left(\frac{h_{ij}}{H_i} - 1\right) + \beta_2\left(\left(\frac{h_{ij}}{H_i}\right)^2 - 1\right) \tag{3.17}$$

Parameters β_1 and β_2 in Eq. 3.17 were estimated using least-squares methods without an intercept term. The corresponding merchantable volume equation is

$$V_{iu} = \frac{\pi \text{dbh}_i^2 H_u}{40000}\left[\frac{1}{6H_i^2}(3\beta_1 H_u H_i + 2\beta_2 H_u^2) - (\beta_1 + \beta_2)\right] \tag{3.18}$$

When H_u equals H_i, the merchantable volume equals the total tree volume.

Trincado and Burkhart (2006) improved the segmented equation initially proposed by Max and Burkhart (1976) using non-linear mixed-effect model techniques to account for within and between individual variations. They also proposed an approximate Bayesian estimator for the mixed parameters that permits calibration of the stem profile for individual trees given two extra upper stem diameter measurements. The segmented equation is

$$\left(\frac{dob_{ij}}{\text{dbh}_i}\right)^2 = \beta_1(q_{ij} - 1) + \beta_2(q_{ij}^2 - 1) + \beta_3(\alpha_2 - q_{ij})^2 I_1 + \beta_4(\alpha_1 - q_{ij})^2 I_2 \tag{3.19}$$
$$I_s = 1 \,\text{if}\, (\alpha_s - q_{ij}) \geq 0,\ 0 \,\text{otherwise}\, (s = 1, 2)$$

We used the R-package *nlme* (Pinheiro et al., 2018) to fit Eq. 3.19. We tested improvements in the fitting by considering non-constant variance on residuals or presence of autoregressive error structure. To include the calibration routine and compare their performance with P-Spline, two additional upper stem diameters for each tree in our sample data were found by interpolation using the nearest measurements at heights of 3.7 m and 6.1 m.

We called it Eq. 3.19r to differentiate from the equations without calibration Eq. 3.19. The respective merchantable volume equation is

$$\begin{aligned} V_{iu} &= \frac{\pi \mathrm{dbh}_i^2 H_u}{40000} \left\{ \frac{\beta_2}{3} H_u^3 + \frac{\beta 1}{2} H_u^2 - (\beta_1 + \beta_2) H_u - \right. \\ &\left. \frac{\beta_3}{3} \left[(\alpha_1 - H_u)^3 I_1 - \alpha_1^3 \right] - \frac{\beta_4}{3} \left[(\alpha_2 - H_u)^3 I_2 - \alpha_2^3 \right] \right\} \end{aligned} \tag{3.20}$$

Fang et al. (2000) developed a compatible taper model system where the taper is a three-segmented step function based on a variable form differential equation proposed initially by Fang and Bailey (1999).

The taper equation is

$$dob_{ij} = c_1 \sqrt{H_i^{(k-\beta_1)/\beta_1} (1 - q_{ij})^{(k-\beta)/\beta} p_1^{I_1 + I_2} p_2^{I_2}} \tag{3.21}$$

where

$$\begin{aligned} c_1 &= \sqrt{V(\mathrm{dbh}_i, H_i) H_i^{k/\beta_1} / [\beta_1 (t_0 - t_1) + \beta_2 (t_1 - p_1 t_2) + \beta_3 p_1 t_2]} \\ \beta &= \beta_1^{1-(I_1+I_2)} \beta_2^{I_1} \beta_3^{I_2}, \\ t_0 &= (1 - h_0/H_i)^{k/\beta_1}, \\ t_1 &= (1 - \alpha_1)^{k/\beta_1}, \\ t_2 &= (1 - \alpha_2)^{k/\beta_2}, \\ p_1 &= (1 - \alpha_1)^{(\beta_2 - \beta_1) k/\beta_1 \beta_2}, \\ p_2 &= (1 - \alpha_2)^{(\beta_3 - \beta_2) k/\beta_2 \beta_3}, \\ k &= \pi/40000, \\ I_1 &= 1 \text{ if } \alpha_1 \leq h_{ij}/H_i \leq \alpha_2, \ 0 \text{ otherwise}, \\ I_2 &= 1 \text{ if } \alpha_2 \leq h_{ij}/H_i \leq 1, \ 0 \text{ otherwise}, \end{aligned}$$

β_1, β_2, β_3, α_1, and α_2 are parameters to be estimated. α_1 and α_2 are the inflection points for the three-segmented equations, and h_0 is the height at the lowest diameter measurement. $V(\mathrm{dbh}_i, H_i) = \beta_4 \mathrm{dbh}_i^{\beta_5} H_i^{\beta_6}$ is the total volume equation, which is one of the three interdependent components included in the compatible system. The volume parameters β_4, β_5, and β_6 were fitted independently of the taper equation. After substitution of volume parameters, the system's remaining five parameters were estimated using non-linear least-squares methods. The corresponding merchantable volume equation is

$$\begin{aligned} V_{iu} &= c_1^2 H_i^{k/\beta_1} [\beta_1 t_0 + (I_1 + I_2)(\beta_2 - \beta_1) t_1 + \\ &\quad I_2(\beta_3 - \beta_2)\alpha_1 t_2 - \beta(1 - q_{ij})^{k/\beta} \alpha_1^{I_1+I_2} \alpha_2^{I_2}] \end{aligned} \tag{3.22}$$

Kozak (2004) described two variable-exponent models called "01" and "02". Equation "01" is

$$dob_{ij} = \beta_1 \mathrm{dbh}_i^{\beta_2} X_{ij}^{\beta_3 + \beta_4 [\frac{1}{e^{\mathrm{dbh}_i/H_i}}] + \beta_5 \mathrm{dbh}_i^{X_i} + \beta_6 X_{ij}^{\mathrm{dbh}_i/H_i}} \tag{3.23}$$

where $X_{ij} = \left[\frac{1-(q_{ij}^{1/4})}{1-p^{1/4}}\right]$, $p = 0.01$. The variable-exponent equation "02" is

$$dob_{ij} = \beta_1 \mathrm{dbh}_i^{\beta_2} H_i^{\beta_3} X_{ij}^{\beta_4 q_{ij}^4 + \beta_5 [\frac{1}{e^{\mathrm{dbh}_i/H_i}}] + \beta_6 X_{ij}^{0.1} + \beta_7 [\frac{1}{\mathrm{dbh}_i}] + \beta_8 H_i^{Q_{ij}} + \beta_9 X_{ij}} \tag{3.24}$$

where

$$X_{ij} = \left[\frac{1 - q_{ij}^{1/3}}{1 - p_i^{1/3}}\right], \qquad Q_{ij} = [1 - q_{ij}^{1/3}],$$
$$p_i = \frac{1.37}{H_i}$$

We used a log-transformation to linearizing these equations. Then, the parameters $\beta_1, \ldots,$ β_9 were estimated using ordinary least-squares. Bias can be introduced by logarithmic transformation (Baskerville, 1972). We corrected potential bias when half of the mean residual sum of squares was greater than 1%. Merchantable volume for models "01" and "02" is calculated using numerical integration.

3.2.8 Performance Evaluation

Leave-one-out cross-validation methods were used to assess the stem diameters prediction error and volume prediction error for the six parametric models and the P-Spline. Initially, one tree is left out, and each model was fitted with the remaining data set. Then, the diameters over bark for the selected tree and the volume in each section were estimated using each model and compared with the observed diameter over bark and volume section, respectively.

The bias for diameter outside bark predictions (*dob*-bias) on the ith tree at the q_{ij}th relative height was defined as $d_{ij} - \hat{dr}_{(-i)jl} * \text{dbh}_i$. The bias in volume was found comparing the Smalin's bole-volume and the respective estimation using the merchantable equation for each model. To evaluate the performance of the models, we analyze both the bias and $RMSD$ for all tree measurements within each of 10 relative height classes along the stem (say $0 < q_{ij} \leq 0.1, 0.1 < q_{ij} \leq 0.2, \ldots, 0.9 < q_{ij} \leq 1$). We consider these ten groups to be adequate to judge the models' prediction performance along the tree stem.

An estimate of the model's precision was computed as the sample variance of bias $var(d_{ij} - \hat{dr}_{(-i)jl} * \text{dbh}_i)$, which was used to find bias confidence intervals.

To report the performance results for volume predictions, we introduce the estimated $RMSD$-efficiency of a given taper equation relative to P-Spline Eq. 3.3. That is

$$RMSD - \text{efficiency} = \frac{RMSD \text{ P-Spline}}{RMSD \text{ of a taper equation}} \times 100 \tag{3.25}$$

If a parametric taper equation is equally good or better than P-Spline in a relative height section, their $RMSD$-efficiency will be equal or larger than 100%.

The proposed P-Spline model uses relative diameter outside bark as the dependent variable $dr_{ij} = dob_{ij}/\text{dbh}_i$, where dbh_i is considered fixed and known. However, some of the parametric taper equations we used to compare have square transformations in their dependent variable. For example, Sharma and Oderwald (2001) uses square diameter outside bark dob_{ij}^2, Kozak et al. (1969) and Max and Burkhart (1976) use relative square diameter outside bark $(dob_{ij}/\text{dbh})^2$. If the square-root retransformation is applied to the unbiased estimates of square diameters, then biased estimations of diameters are obtained. We include in those cases the bias-correction proposed by Czaplewski and Bruce (1990).

3.3 Results and Discussion

3.3.1 Parametric Models

Table 3.2 presents the estimated parameters for the single equations, segmented, and variable-exponent equations using all sampled trees. The system of equations of Sharma and Oderwald (2001) (Eq. 3.14 and 3.15) comprises two parameters, but the taper equation only requires β_1. Numerical consistency between the volume and taper equation was obtained by doing a simultaneous estimation of β_1 in both equations. We followed the same strategy reported by Sharma and Oderwald (2001) to fit the system. That is, for each measurement within a tree, the dbh_i, H_i, and V_i were replicated, thus, equal weight was given to each observation.

We fitted the single equation of Kozak et al. (1969) (Eq. 3.17) using standard linear regression without an intercept. Residual diagnostic plots for fitted models from Eq. 3.14 and Eq. 3.17 (see Figure S1 and Figure S2) showed a severe dispersion of residuals at lower sections of the trees, and present an evident lack-of-fit pattern. Consequently, we expect to

see the largest systematic bias on diameter predictions along the stem if those single fitted models are used to predict diameters. Plots for diagnostics of residuals for all the models and plots comparing observed versus predicted relative *dob* for each model are available as supplementary material.

In the Max and Burkhart (1976) taper equation (Eq. 3.19), the parameters α_1 and α_2 indicate the location of knots for the transition between three quadratic submodels, so that the entire taper is explained for three linked fixed forms: one for the lower section that accounts for any butt swell, the middle model should resemble a paraboloid and the last model in the upper part of the stem must follow a truncated cone. They found values for α_1 and α_2 of 74% and 11% (inside bark taper) of total height, respectively. Here we found that outside bark values are located further apart at 79% and 5.3% of the total height (see Table 3.2).

The best fit for Eq. 3.19 was found using random effects on β_1, β_2, and β_3, and including an exponential function structure as $exp(2 * \theta_1 * h_{ij}/H)$ to account for within-tree variability. The residual diagnostic plot for the Max and Burkhart (1976) taper equation (see Figure S3) shows a more stable and symmetric pattern than the single equations.

The parameters for the volume equation of Fang et al. (2000) (Eq. 3.21) were estimated independently of the taper equation. The resulting model equation is

$$V_i = 3.016643 \times 10^{-05}(\mathrm{dbh}_i^{2.027595})(H_i^{1.037887}) \tag{3.26}$$

$$\mathrm{N} = 144, \text{ Residual standard error} = 0.03967$$

The location of estimated knot values for Eq. 3.21 are 64.6% and 4.3%, resulting in different and lower values compared with Eq. 3.19 (see Table 3.2). It suggests that model assumptions, particularly those related to stem forms on each segment, influence the location of estimated knots. An analysis of studentized residuals for models in Eq. 3.19 and Eq. 3.21

(see Figure S3 and Figure S4) showed that these models had more homogeneous residuals variance than single equations. However, even with this three-segmented equation, large residual variance in lowest stem measurements remains.

Contrary to the segmented models, variable-exponent models try to explain multiple forms in a unique equation. The parameter estimates for the two selected variable-exponent equations from Kozak (2004) (Eq. 3.23 and Eq. 3.24) are presented in Table 3.2. These two models were linearized using logarithmic transformation, and therefore, estimation was more easily achieved using linear least-squares regression. Similar to the single equations, the analysis of studentized residuals showed some evidence of lack-of-fit, principally for the model in Eq 3.23 (see Figure S5 and S6).

3.3.2 Number of Knots for P-Spline and Parameter Estimation

Figure 1 shows the average $RMSD$ outside bark with their standard error bars for a range of numbers of knots $K \in S$ using the P-Spline Eq. 3.3. For this particular sample of trees, the lower $RMSD$ was reached with eight knots ($RMSD = 0.989$); after eight knots, we did not find any further reduction. The P-Spline Eq. 3.2 also reached their minimum at eight knots ($RMSD = 0.990$). In general, the larger the number of spline function basis used with P-Spline, the better the smoothness achieved. However, taper curves are almost monotonic, and with decreasing diameter from the stump to top height, it is reasonable that a smaller number of knots are required.

For both P-Spline models (Eq. 3.2 and Eq. 3.3), we used eight knots spaced at percentile positions in the relative range of heights. Consequently, P-Spline Eq. 3.2 has 12 parameters, and P-Spline Eq. 3.3 has 21 parameters. We recommend this cross-validation method to choose an adequate number of knots if this model is probed with a more extensive data set or data from another species.

No improvements in fitting were gained when including a general structure for within-individual correlations and non-homogeneous variance in the variance-covariance matrix of the error term. Therefore, a more simple structure was used such as $\mathbf{R}_i = \sigma^2_\epsilon \mathbf{I}_{n_i}$ where $\mathbf{I}_{n_i}$ is an identity matrix $(n_i \times n_i)$. Also we assumed that $\mathbf{G} = \sigma^2_{\mathbf{u}} \mathbf{I}$.

An important feature of the model presented here is its link to linear mixed-effect models. The fitted parameters in P-Spline are the Best Linear Unbiased Prediction (BLUP) in a particular mixed-effect model. That is, for arbitrary N $\times$ 1 vectors $\mathbf{s}$ and $\mathbf{t}$, the estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$ minimize the prediction error $E\{(\mathbf{s}^T\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{t}^T\mathbf{Z}\hat{\mathbf{u}}) - (\mathbf{s}^T\mathbf{X}\boldsymbol{\beta} + \mathbf{t}^T\mathbf{Z}\mathbf{u})\}^2$ subject to the unbiasedness condition $E(\mathbf{s}^T\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{t}^T\mathbf{Z}\hat{\mathbf{u}}) = E(\mathbf{s}^T\mathbf{X}\boldsymbol{\beta} + \mathbf{t}^T\mathbf{Z}\mathbf{u})$ (Robinson, 1991), and specifically, the vector of spline basis parameters $\beta_{3+k} \sim N(0, \sigma^2_{\beta_{3+k}} D)$, where $D = I$ and $\sigma^2_{\beta_{3+k}} = \sigma^2_\epsilon / \lambda$. Using all the available data with Eq. 3.3, we found that $\hat{\sigma}^2_\epsilon = 0.001851$ and $\hat{\sigma}^2_{\mathbf{u}} = 191725.7$. Therefore, the smoothing parameter takes a value close to zero, indicating that our model has imposed small penalization on the parameters of the selected cubic spline basis with this particular data set. In other words, we can assimilate the results of our model as an eight-segmented equation, but this may not be the case with other tree species or data. The big value for $\hat{\sigma}^2_{\mathbf{u}}$ is also a consequence of the considerable observed variability on estimated random components (see Table 3.2). As the RML method assures an optimal amount of smoothing, smaller λ values lead to a decrease in bias but an increase in residual variance.

Table 3.3 show the estimate parameters for Eq. 3.2 and Eq. 3.3. The log-Likelihood with Eq. 3.3 was 5268.3, which is only 0.4% larger than the log-Likelihood with the P-Spline without dbh-class (Eq. 3.2). However, the approximate Chi-Square test returns a p-value < 0.0001, indicating an important contribution of the dbh-class as an additive variable in the model.

Note that there is a substantial difference between our model and the model presented by Kublin et al. (2013). First, they used a B-spline basis to model the population mean

taper (truncated spline basis and B-Spline basis matrices of the same degree are equivalent through the existence of a square invertible matrix, see Eilers and Marx (1996) or Ruppert et al. (2003)). Secondly, Kublin et al. (2013) used an additional B-Spline to model the tree-individual variations from the mean taper, allowing the variance to vary locally. Contrary to the Kublin et al. (2013) model, our model has the limitation that a single parameter, $\sigma^2_{\beta_{3+k}}$, is used to shrink all the spline coefficients. However, this is not a problem because we know that the underlying taper function in loblolly pine does not have complex changes or extreme varying heterogeneity.

Graphical diagnostics of studentized residuals for the P-Spline did not suggest a lack-of-fit and indicated homogenized residual variance (see Figure 2). A scatter-plot of paired observed and predicted diameters along the stem shows an expected behavior from the breast height to the top. Large residual variability was found between the stump and at the dbh-height, i.e., relative heights between 0 and 0.1 in both P-Spline models (see Figure 2). This extra variability can be due to rapid changes in taper close to the stump. These less precise predictions towards the stump were also observed with the parametric models (see plots in the Supplemental Material). Lack of precision at the bottom of a stem can lead to the biased diameter and merchantable volume estimations. We evaluated the magnitude of such bias in the next sections.

3.3.3 Assessing Stem Diameter Predictions

Table 3.4 shows the average cross-validation bias for each model at different relative heights classes. Figure 3 shows the average cross-validation bias as well as the corresponding 95% confidence intervals for each selected model in each relative height class. The confidence intervals were used to support the prevalence of biased or unbiased diameter predictions for each model in each relative height class. Specifically, we analyze the average bias along the stem by checking in which height classes the confidence intervals contain zero, hence it gives

us an evaluation of the compatibility or incompatibility of the model assumptions with the observed data.

The parametric single and segmented taper equations did not produce consistent predictions along the stem (see Figure 3), and significant bias was observed across the relative height classes. For example, the single equation Eq. 3.14 tends to overestimate the predicted diameter over bark near the top section of the tree. It shows underestimations in six of the lower height classes, indicating a poor performance for taper modeling with our sample. Eq. 3.17, Eq. 3.19, and Eq. 3.19r tended to overestimate the diameter at the lower height classes (up to 40% of the total height from the ground), and from the middle to the top height class sections, they tend to produce underestimations.

The segmented model Eq. 3.21 produces overestimation between 10% and 40% of the tree height and performs well from the middle to the top. Even though the estimated parameters for the segmented models are unbiased and the residuals diagnostic does not show lack of fit, constraining the taper to three-segmented forms is still restrictive. We observed important bias along the stem with these segmented models compared with the P-Spline (Table 3 and Figure 3). These results suggest that more than two knots are required to reach a satisfactory taper prediction.

The variable-exponent model Eq. 3.23 is less accurate than the Kozak (2004) model "02" (Eq. 3.24). Eq. 3.23 produces overestimation until 30% of the height, and underestimation around the middle sections of the trees.

High flexibility of the variable exponent model was reported by Sharma and Zhang (2004). The results in Table 3.4 indicate that the P-Spline has comparable or better performance. The two models proposed by Kozak (2004) produce biased predictions of diameters outside bark, predominantly on lower and mid-height logs (until the 60% of tree relative height). These results indicate that the included flexibility in those complex non-linear expressions does not increase the accuracy in diameter estimation as much as the P-Splines does.

The simple P-Spline model (Eq. 3.2) shows precise and unbiased estimations for diameter outside bark across most of the tree relative heights.

Even better than model Eq. 3.19r, which uses two extra upper diameters to calibrate the taper curve. While the addition of the dbh-class factor variable in the simple P-Spline model produces a remarkably accurate gain in diameter prediction, between 10 and 40% of tree height (see Table 3.4), increased biased predictions were observed on the top 20% tree height. However, on average, 62% of total tree volume outside bark is accumulated on the first 40% of tree height; therefore, more reliable taper estimates on lower stem segments should result in better estimates of volume.

The bottom line in Table 3.4 shows the overall *dob*-bias using all the observations in the sample trees. Based on this evaluation, the variable-exponent models and the segmented model of Fang et al. (2000) report unbiased diameter predictions. Evaluating performance at the total tree-level can hide localized bias in sections of the stem. For example, Eq. 3.23 shows over predictions on the lower relative height classes and underpredictions on the middle relative height classes. However, at tree-level, the bias cancels out, resulting in apparently acceptable performance.

3.3.4 Assessing Volume Predictions

Table 3.5 shows the Tree Form Factors (TFF) obtained by the defined integration of the square expected mean diameter in each dbh-class and respective variance when Eq. 3.3 is used to calculate total volume. Decisions about the size and number of dbh-classes to include in the model should depend on the sample size available to fit the model. Here we decided to use ten groups with approximately the same sample size. We did not evaluate the effect of choosing different dbh classes on the quality of our model or their ability to predict volume.

When only total volume estimates are desired, using those TFF could simplify the computation when processing large inventory data sets. For example, a tree with dbh = 25.7

cm and a total height of $H = 22.2$ m, is located in the dbh-class eight: $(25.6 - 27.2]$, which correspond to a $TFF = 0.46955$; the total volume over bark for this tree is $V = \pi * 2.5e^{-5} * 0.46955 * (25.7)^2 * 22.2 = 0.5407$ m^3. While for total volume estimates TFF result beneficial, for the merchantable volume, it is necessary to perform a numerical integration with Eq. 3.10 changing the integration limits accordingly. Figure 4 shows the high correlation ($\rho = 0.994$) between the estimated total tree outside bark volume calculated with TFF and the volume obtained using Smalian's formula.

In terms of average volume-bias, segmented and variable exponent parametric equations perform close to the P-Spline models (see Table S1 in supplementary material). For a more clear comparison in volume, we discuss here the $RMSD$-efficiency (see Table 3.6). The simple P-Spline model produces more accurate volume predictions than P-Spline Eq. 3.3 in the firs three relative height sections (0 - 30% of relative heights). Specifically, the simple P-Spline produces between 5.4% to 9% more $RMSD$-efficiency) in the lower 30% of relative height where on average is located the 48% of the total tree volume.

While the inclusion of a dbh-class factor variable improves the P-Spline capability for unbiased and precise *dob* predictions at the lowest stem sections, that effect does not appear with volume predictions at the same location on the stem (see Table 3.6).

P-Spline Eq. 3.3 consistently performs equal or better than the other parametric equations in the range of 40% to 80% of the relative heights, corresponding to an average of 27.9 % of total volume. The calibration of Max and Burkhart (1976) equation with two extra upper-stem diameter measurements produces an increment in efficiency around 2.6 % to 7.3% compared with P-Spline Eq. 3.3 in the first 40% of relative heights. It is only more efficient than the simple P-Spline between 30% to 50% of the relative heights.

Previous works have shown that taper calibration with extra upper-stem diameters measurements could improve volume estimations (Cao and Wang, 2011; Trincado and Burkhart,

2006; Kublin et al., 2008; Lappi, 2006). Results in Table 3.6 indicate that the simple P-Splines can perform equal to or better than the segmented and calibrated segmented equations.

3.3.5 Confidence Intervals for Volume[3]

Confidence intervals for volume can be obtained with Eq. 3.13. We found that the volume uncertainty increase as the upper integration limit move from the ground to the tree-tip. For example, Figure 3.5 shows the accumulated volume from the ground up to the tip for a single tree with $dbh = 36.65$ cm and $H = 29.69$ m. The dotted line represents the 95% confidence interval (Eq. 3.13).

Statistical inference on tree-volume derived from taper equations has received little attention in the literature. However, it is the appropriate metric to report uncertainty. For example, the confidence intervals presented in Eq. 3.13 can be used to compare individual tree-taper profiles. We believe it is necessary when the effect of cultural treatments or gains on improved genetic material need to be precisely assessed.

3.4 Practical Implementation[4]

(This section is not included in the published contribution). In order to make this contribution suitable, the proposed model was coded in several R functions and compiled in an R package **PSTapeR**. The package contains R functions for fitting a taper equation using penalized spline linear mixed models and visualization. This package includes the following functionalities:

- Calculate confidence intervals for the estimated taper curve.
- Estimate the total and merchantable tree volume.

[3]This sub-section was not included in the submitted paper
[4]This sub-section was not included in the submitted paper

- Compute confidence intervals for total tree volume and merchantable volume.
- Options for deploy parallel computing when processing large inventory jobs.

To install the package use: `install.packages("PSTapeR")`

A detailed description of the functions and examples is included in the package manual. The tree-sections information for each of 147 loblolly pine trees used in this dissertation is included in **PSTapeR**. For industrial and regional applications, it is recommended to calibrate the model with a more representative data set. Given a dataset **PSTaperR** generates diameter-class groups automatically, and these are used as an additive predictor variable.

The command line `data(loblolly_buckdata)` load the stem analysis dataset used in this dissertation. `data(loblolly_pine_trees)` load a table with the tree-level variables.

The function `plot.PSTapeR` produces a plot with the estimated profile for a particular tree and includes the simultaneous confidence interval proposed by Krivobokova et al. (2010). Figure 3.6 show an example of the estimate profile for one tree with dbh = 10.2 in and $H = 72.35$ft with its respective 95% simultaneous confidence interval.

The package also allows uncertainty assessments for volume. Confidence intervals for total tree volume and merchantable volume are available in the output of `PSTapeR.Predict.Volume` using Eq. 3.13. Confidence intervals for volume require extra computational resources. Figure **??** shows the elapsed time to process a tree-list of size 10,000 using **PSTapeR**. When the simulation includes confidence intervals, the total computation time dramatically increases, on average, 0.02 seconds per tree. The option for parallel computing helps reduce the computing time for merchantable volume with confidence intervals, reducing the total waiting time for 17.8 percent. The package can scale horizontally, so depending on the size of computing resources available, the waiting time for the case of merchantable volume can be reduced significantly.

3.5 Conclusions

A semiparametric approach was proposed to model the tree stem form of loblolly pine for the non-thinned plantations at the Whitehall Forest at the University of Georgia. The model is a penalized spline regression, P-Spline, of degree $p = 3$. We also tested the performance of an extended P-Spline with an additive dbh-class factor variable. The proposed extended P-Spline method outperforms the traditional parametric taper equations when used to predict outside bark diameters in the lower portions of the stem, up to 40% of the tree relative height where the more valuable wood products (on average 62% of the total outside bark volume) are contained.

The Fang et al. (2000) taper equations showed the best performance to predict diameters outside bark on the top stem sections (40-60% and 70-90%).

While the inclusion of a dbh-class variable improves the P-Spline fitting and their capability for explaining taper shapes, that comparative advantage was not reflected in a more accurate ability in terms of volume predictions. Instead, the simple P-Spline showed the best performance in terms of volume. However, the proposed extended P-Spline performs similar to the best parametric equation by Max and Burkhart (1976) with taper calibration.

Although taper calibration with extra upper diameter measurements has proved to increase volume prediction accuracy (Trincado and Burkhart, 2006; Cao and Wang, 2011), the simple P-Spline model presented here shows superior performance without any additional calibration. Consequently, the use of semiparametric taper modeling could result in savings on inventory costs by omitting any additional measurements.

Our comparison based on cross-validations shows that P-Spline portrays a flexible modeling framework with low bias and allowing precise diameter and volume predictions along the stem. Consequently, the *a priori* fixed forms assumed by the empirical taper equations imposes an unnecessary restriction that fails to explain the tree form adequately when com-

pared with P-Spline. Further, we expect that the use of P-Spline for loblolly pine volume estimations will be more reliable because minimize error due to model misspecifications.

The proposed semiparametric modeling framework can be explored with a more geographically extended dataset to obtain a representative taper and volume equation for loblolly pine. We recommend using a cross-validation technique to determine the adequate number of knots and to decide the number of dbh-classes.

Table 3.1: Summary statistics for 147 loblolly pine (*Pinus taeda* L.) trees selected from Whitehall Forest and used to fit the taper models, where dbh = diameter at breast height (1.37 m) in cm, H = total tree height in m, and V = volume in cubic meters.

Variable	Mean	Min	Max	Standard deviation
dbh (cm)	23.29	13.59	46.91	5.68
H (m)	22.45	15.69	31.46	3.72
V (m^3)	0.509	0.110	2.55	0.362

Table 3.2: Estimated parameters (Standard Error) for six parametric taper equations.

	Model					
Parameters	Eq. 3.14	Eq. 3.17	Eq. 3.19	Eq. 3.21	Eq. 3.23	Eq. 3.24
$\hat{\beta}_1$	2.0599	-2.1226	-4.4999	5.074314×10^{-6}	-0.151041	-0.246359
	(0.00237)	(0.02711)	(0.44270)	(1.405×10^{-7})	(0.030822)	(0.044944)
$\hat{\beta}_2$	0.4909	0.8557	2.1471	3.733×10^{-5}	1.108733	1.079371
	(0.00521)	(0.02221)	(0.24505)	(1.859×10^{-7})	(0.009735)	(0.014373)
$\hat{\beta}_3$			-2.1301	3.054×10^{-5}	0.371550	0.003502
			(0.23906)	(3.956×10^{-7})	(0.012031)	(0.020608)
$\hat{\beta}_4$			290.323	3.017×10^{-5}	0.182456	0.335497
			(14.14348)	(3.464×10^{-6})	(0.032547)	(0.016581)
$\hat{\beta}_5$				2.028	0.054284	-0.408082
				(3.424×10^{-2})	(0.002861)	(0.050967)
$\hat{\beta}_6$				1.038	-0.579400	0.473529
				(5.481×10^{-2})	(0.028073)	(0.018226)
$\hat{\beta}_7$						1.410832
						(0.321457)
$\hat{\beta}_8$						0.062198
						(0.005860)
$\hat{\beta}_9$						-0.305270
						(0.046171)
$\hat{\alpha}_1$			0.0531	0.0431		
			(0.00118)	(0.00101)		
$\hat{\alpha}_2$			0.7906	0.6463		
			(0.01567)	(0.01157)		

Table 3.3: Estimated fixed-effect parameters (Standard Error) and estimated random-effect (Standard Deviation) for the simple P-Spline taper equation (Eq. 3.2) and the P-Spline including the factor (α_m, $m = 2, \ldots, 10$) dbh-class as an additive variable (Eq. 3.3).

Fixed effect	Estimate (SE) Eq. 3.2	Estimate (SE) Eq. 3.3	Random effect	Estimate (SD) Eq. 3.2	Estimate (SD) Eq. 3.3
$\hat{\beta}_0$	0.7143 (0.0009)	0.6213 (0.0027)	$\hat{\beta}_4$	2145.609 (763.2419)	1234.463 (437.8649)
$\hat{\beta}_1$	-17.6374 (0.8740)	-15.4359 (0.5655)	$\hat{\beta}_5$	70.0287 (763.2419)	23.4410 (437.8649)
$\hat{\beta}_2$	318.2402 (25.1002)	228.7192 (12.6601)	$\hat{\beta}_6$	3.3206 (763.2419)	0.0248 (437.8649)
$\hat{\beta}_3$	-2218.103 (207.1810)	-1258.200 (80.8364)	$\hat{\beta}_7$	-3.1995 (763.2419)	-1.5008 (437.8649)
$\hat{\alpha}_2$		0.0251 (0.0038)	$\hat{\beta}_8$	2.9280 (763.2419)	-0.7846 (437.8649)
$\hat{\alpha}_3$		0.0004 (0.0038)	$\hat{\beta}_9$	-6.7217 (763.2419)	4.4877 (437.8649)
$\hat{\alpha}_4$		0.0240 (0.0037)	$\hat{\beta}_{10}$	18.9196 (763.2419)	-2.9301 (437.8649)
$\hat{\alpha}_5$		0.0159 (0.0037)	$\hat{\beta}_{11}$	-36.0333 (763.2419)	16.7436 (437.8649)
$\hat{\alpha}_6$		0.0135 (0.0037)			
$\hat{\alpha}_7$		0.0154 (0.0036)			
$\hat{\alpha}_8$		0.0103 (0.0037)			
$\hat{\alpha}_9$		0.0222 (0.0035)			
$\hat{\alpha}_{10}$		0.0271 (0.0035)			

Table 3.4: Cross-validation average bias on diameter outside bark predictions by relative height classes with six parametric taper equations and two P-Spline (Eq. 3.2 and Eq. 3.3). The lowest bias in each relative height class is bolded.

Relative height	No. obs.	Model								
		Eq. 3.2	Eq. 3.3	Eq.3.14	Eq. 3.17	Eq. 3.19	Eq. 3.19r	Eq. 3.21	Eq. 3.23	Eq. 3.24
		Mean Bias								
(0-0.1]	485	**-0.033**	-0.039	-3.148*	-1.057*	-0.959*	-0.948*	0.196*	-0.228*	-0.187*
(0.1,0.2]	283	-0.015	**-0.007**	0.285*	-1.741*	-0.338*	-0.317*	-0.168*	-0.574*	-0.105*
(0.2-0.3]	268	0.037	**0.025**	0.774*	-0.964*	-0.299*	-0.270*	-0.189*	-0.235*	0.159*
(0.3-0.4]	268	0.042	**0.034**	0.918*	-0.215*	-0.191*	-0.157*	-0.155*	0.095	0.153*
(0.4-0.5]	268	0.115	0.121	1.037*	0.561*	0.073	0.109	**0.000**	0.528*	0.221*
(0.5-0.6]	271	0.100	0.079	0.909*	1.064*	0.217*	0.259*	**-0.025**	0.728*	0.196*
(0.6-0.7]	268	**0.047**	0.072	0.484*	1.213*	0.247*	0.290*	-0.133	0.618*	0.141
(0.7-0.8]	260	0.065	0.150	-0.335*	0.826*	0.120	0.161*	**0.008**	0.051	-0.019
(0.8-0.9]	144	-0.227	0.383*	-1.170*	0.227*	0.224*	0.276*	**0.051**	-0.793*	-0.075
(0.9-1]	148	**0.005**	-0.057*	-0.012	0.010	0.020	0.022	0.018	-0.009	0.022
All	2663	**0.021**	0.058*	-0.224*	-0.116*	-0.181*	-0.151*	-0.033	0.034	0.038

* Denotes low compatibility of the data with the null assumption in a paired t-test between observed diameters and predicted diameters using cross-validation.

Table 3.5: Loblolly pine Tree Form Factors (TFF) by dbh-class to be used in the calculation of total volume outside bark with the equation $V = \pi 2.5e^{-5} TFF \text{dbh}^2 H$ (dbh in centimeters and H in meters).

dbh-class	dbh (cm)	TFF
1	(13.5-16.6]	0.45686
2	(16.6-18.6]	0.48821
3	(18.6-20.1]	0.45737
4	(20.1-21.5]	0.48681
5	(21.5-22.4]	0.47657
6	(22.4-23.9]	0.47361
7	(23.9-25.6]	0.47588
8	(25.6-27.2]	0.46955
9	(27.2-30.4]	0.48445
10	(30.4-47]	0.49072

Table 3.6: Cross-validation $RMSD$-efficiency (%) relative to P-Spline Eq. 3.3. The six parametric taper equations and the simple P-Spline Eq. 3.2 are compared with Eq. 3.3 using the definition in Eq. 3.25.

Relative height	No. obs.	Model								
		Eq. 3.2	Eq. 3.3	Eq.3.14	Eq. 3.17	Eq. 3.19	Eq. 3.19r	Eq. 3.21	Eq. 3.23	Eq. 3.24
		$RMSD$- efficiency (%)								
(0-0.1]	485	**109.031**	100	75.338	48.657	107.261	107.142	97.520	98.520	85.789
(0.1,0.2]	283	**118.355**	100	85.322	36.579	101.446	106.696	90.868	91.817	99.622
(0.2-0.3]	268	**105.479**	100	94.828	54.507	97.582	103.876	93.766	89.670	103.666
(0.3-0.4]	268	93.313	100	73.793	92.319	98.285	**102.584**	98.805	81.028	89.260
(0.4-0.5]	268	94.023	**100**	69.229	81.826	96.553	98.626	99.086	76.500	90.188
(0.5-0.6]	271	90.746	**100**	68.092	61.218	87.942	88.048	99.018	68.561	86.799
(0.6-0.7]	268	93.663	100	79.097	56.995	87.463	87.147	**101.482**	73.813	89.779
(0.7-0.8]	260	96.023	100	99.437	67.883	94.394	93.505	**100.695**	96.269	98.604
(0.8-0.9]	144	103.039	100	73.811	96.489	103.561	103.125	103.163	86.171	**106.462**
(0.9-1]	148	57.773	**100**	31.100	67.524	75.044	76.774	71.247	38.539	64.831
All	2663	96.653	**100**	71.960	55.444	95.414	97.382	95.939	78.157	91.018

Table 3.7: Cross-validation $RMSD$-efficiency (%) relative to P-Spline Eq. 3.3. The six parametric taper equations and the simple P-Spline Eq. 3.2 are compared with Eq. 3.3 using the definition in Eq. 3.25

Relative height	No. obs.	Model						
		P-Spline1	P-Spline2	M&B(1979)	M&Br(1979)	Fang(1999)	Kozak01(2004)	Kozak02(2004)
		$RMSD$- efficiency (%)						
(0-0.1]	485	**109.031**	100	107.261	107.142	97.520	98.520	85.789
(0.1,0.2]	283	**118.355**	100	101.446	106.696	90.868	91.817	99.622
(0.2-0.3]	268	**105.479**	100	97.582	103.876	93.766	89.670	103.666
(0.3-0.4]	268	93.313	100	98.285	**102.584**	98.805	81.028	89.260
(0.4-0.5]	268	94.023	**100**	96.553	98.626	99.086	76.500	90.188
(0.5-0.6]	271	90.746	**100**	87.942	88.048	99.018	68.561	86.799
(0.6-0.7]	268	93.663	100	87.463	87.147	**101.482**	73.813	89.779
(0.7-0.8]	260	96.023	100	94.394	93.505	**100.695**	96.269	98.604
(0.8-0.9]	144	103.039	100	103.561	103.125	103.163	86.171	**106.462**
(0.9-1]	148	57.773	**100**	75.044	76.774	71.247	38.539	64.831
All	2663	96.653	**100**	95.414	97.382	95.939	78.157	91.018

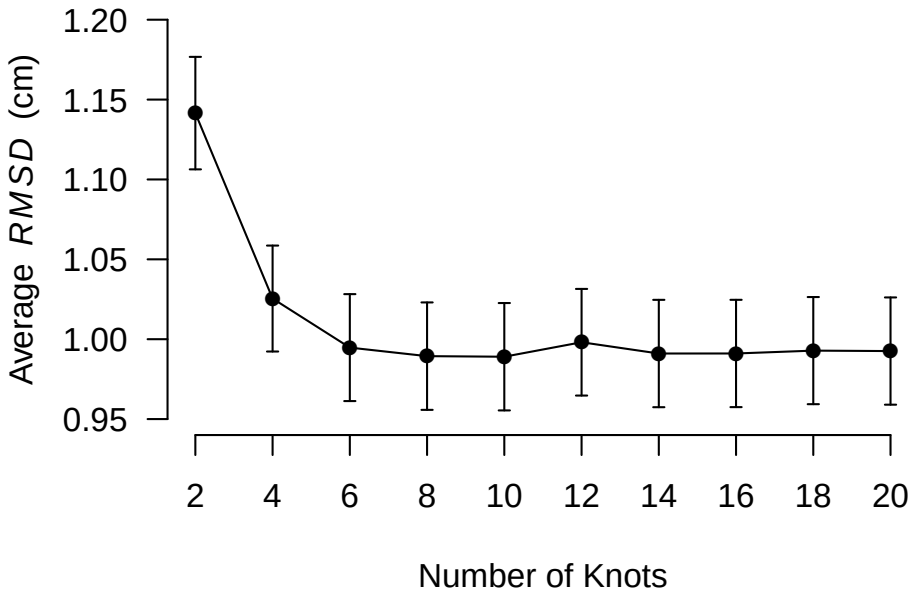

Figure 3.1: Cross-validation average root mean square difference for diameter prediction ($RMSD$) $\pm$ SE using P-Spline taper regression Eq. 3.3 with a different number of knots. Lowest $RMSD$ was achieved using 8 knots ($RMSD = 0.989$; $SE = 0.034$).

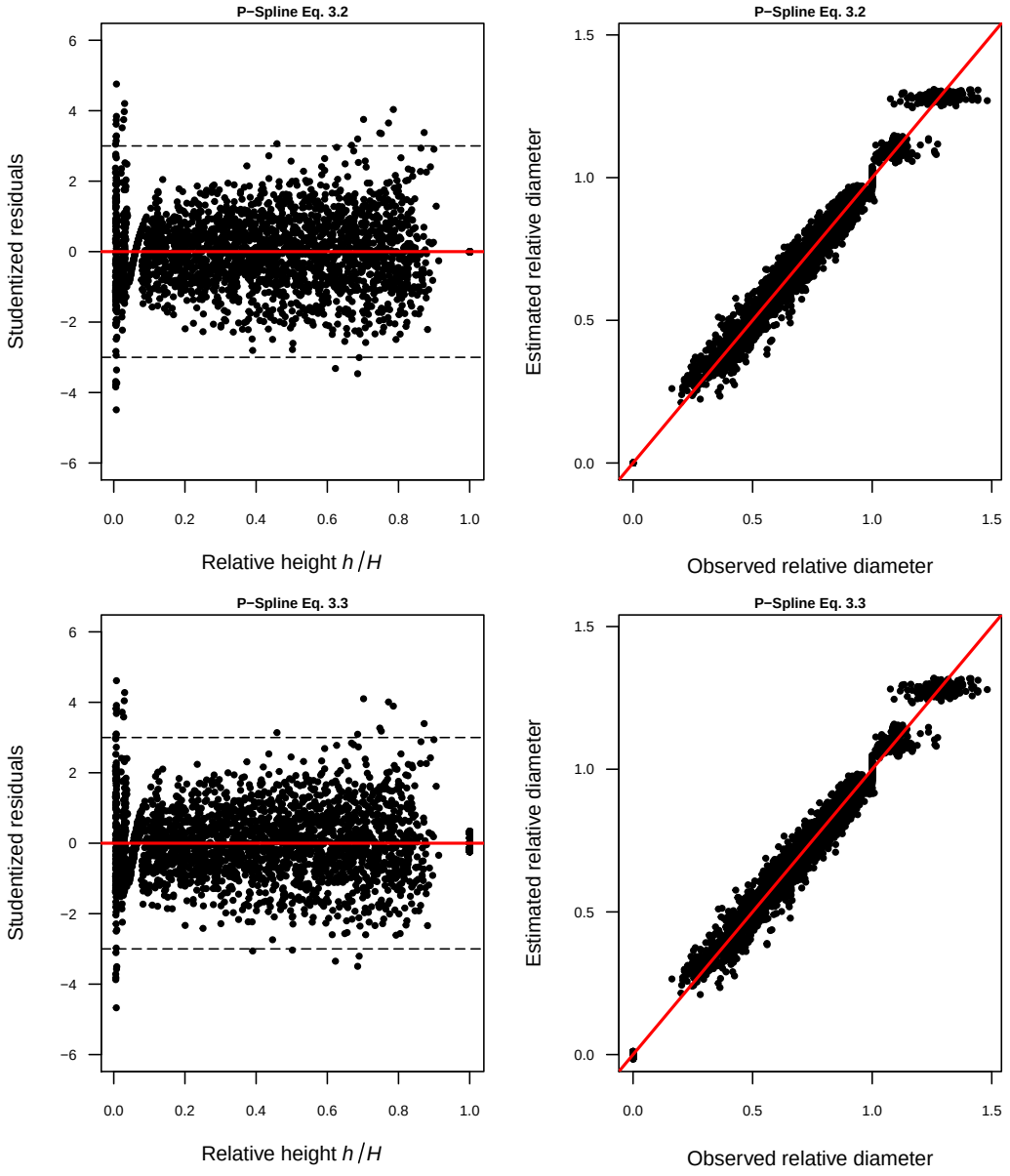

Figure 3.2: Studentized residuals and comparison of predicted and observed relative diameters dob_{ij}/dbh_i for the simple P-Spline taper regression (Eq. 3.2) and P-Spline with dbh-class as an additive factor variable (Eq. 3.3).

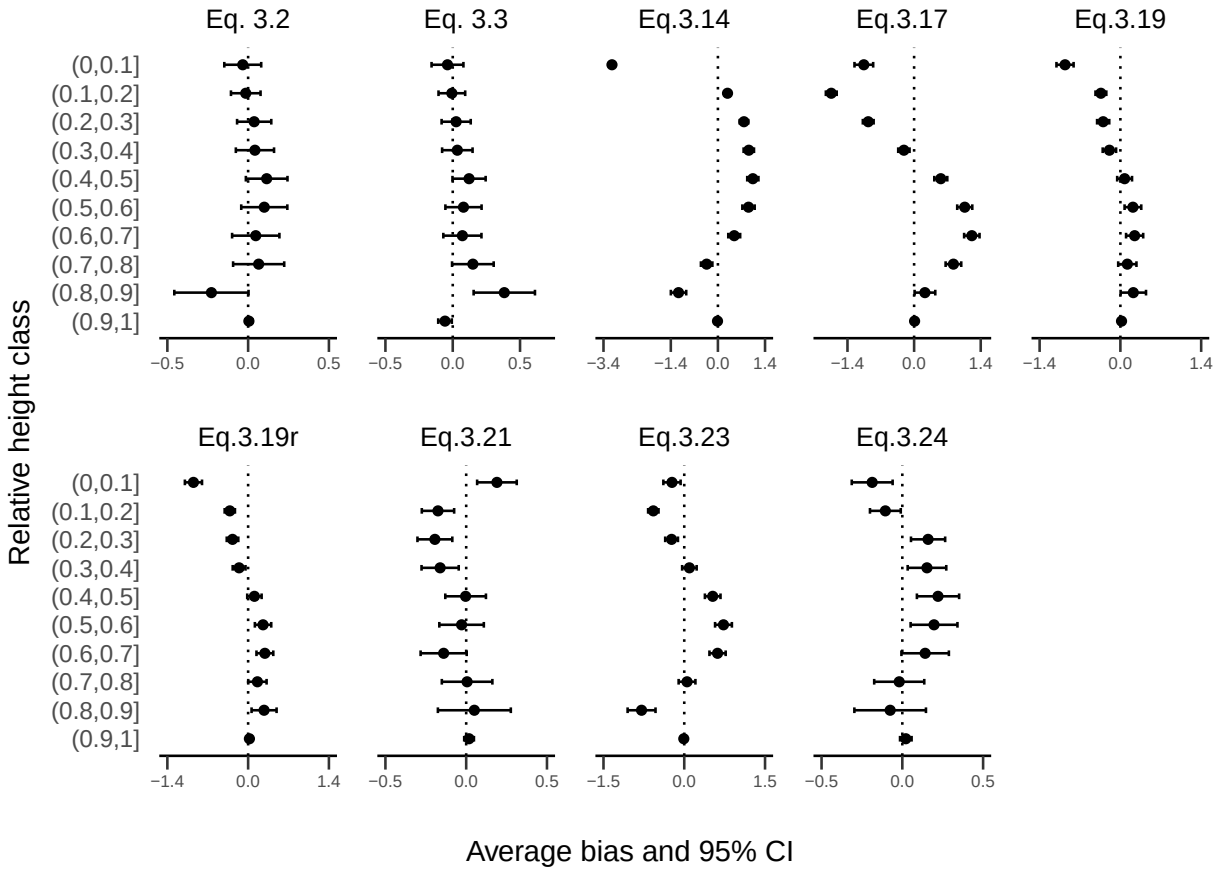

Figure 3.3: Average bias by relative height class along the stem for six parametric taper equations and P-Spline (Eq. 3.2 and Eq. 3.3). The bars represent 95% confidence intervals. Note the scale on the x-axis changes for each equation to allow representation on one figure.

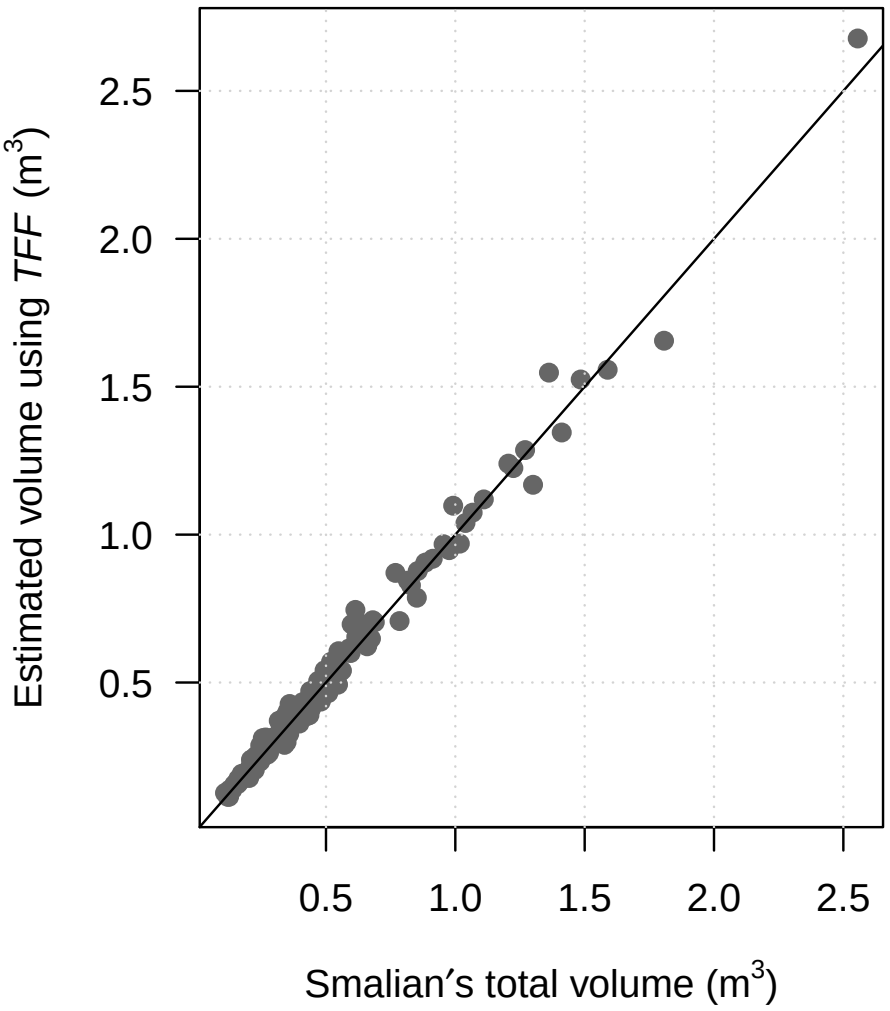

Figure 3.4: Comparison of Smalian's total outside bark volume and predicted volume using TFF (see Table 3.5) for the 147 loblolly pine sampled trees. The line has intercept zero and slope one.

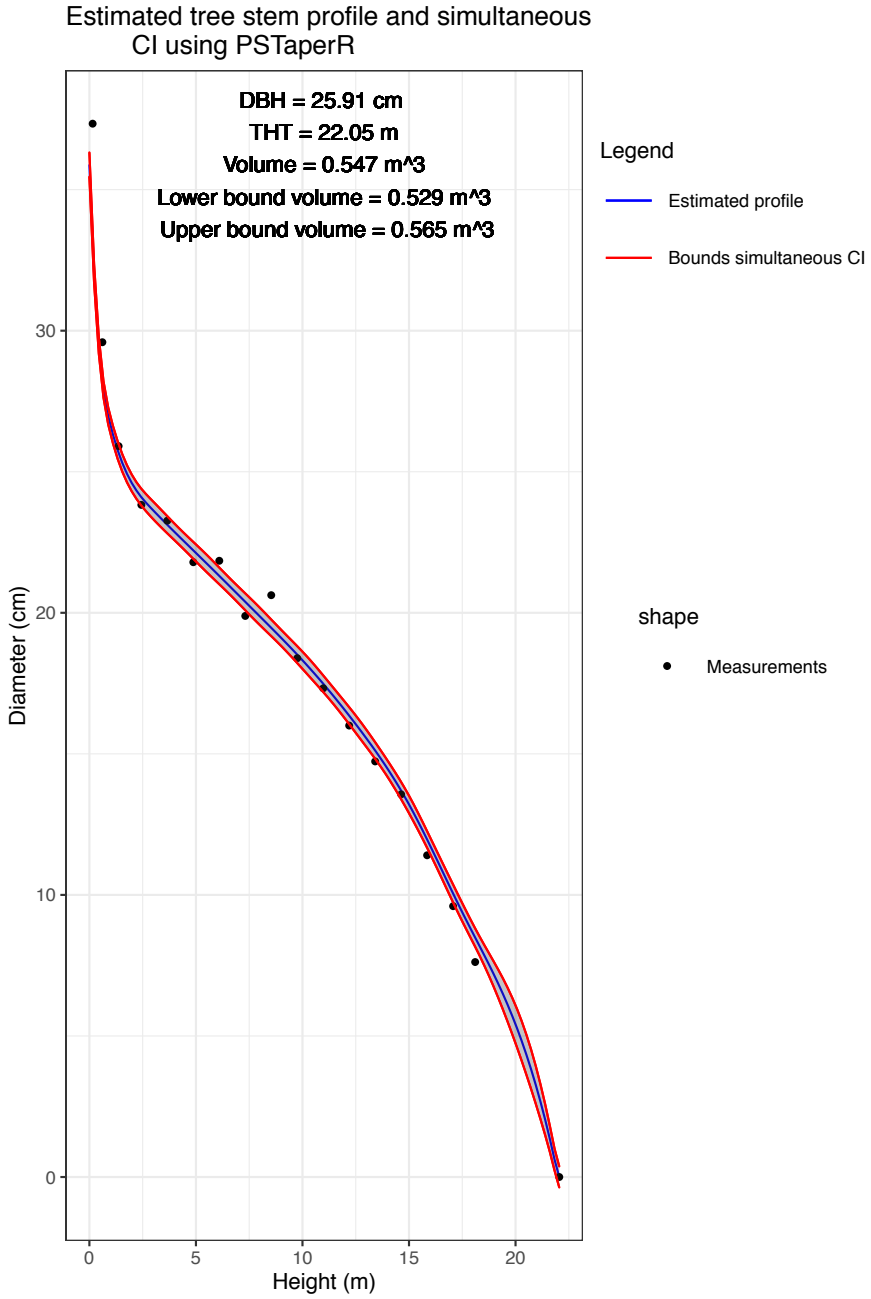

Figure 3.5: Observed diameters and estimate profile for one tree with dbh = 10.2 in and $H = 72.35$ft. The shaded area represents a 95% simultaneous confidence interval.

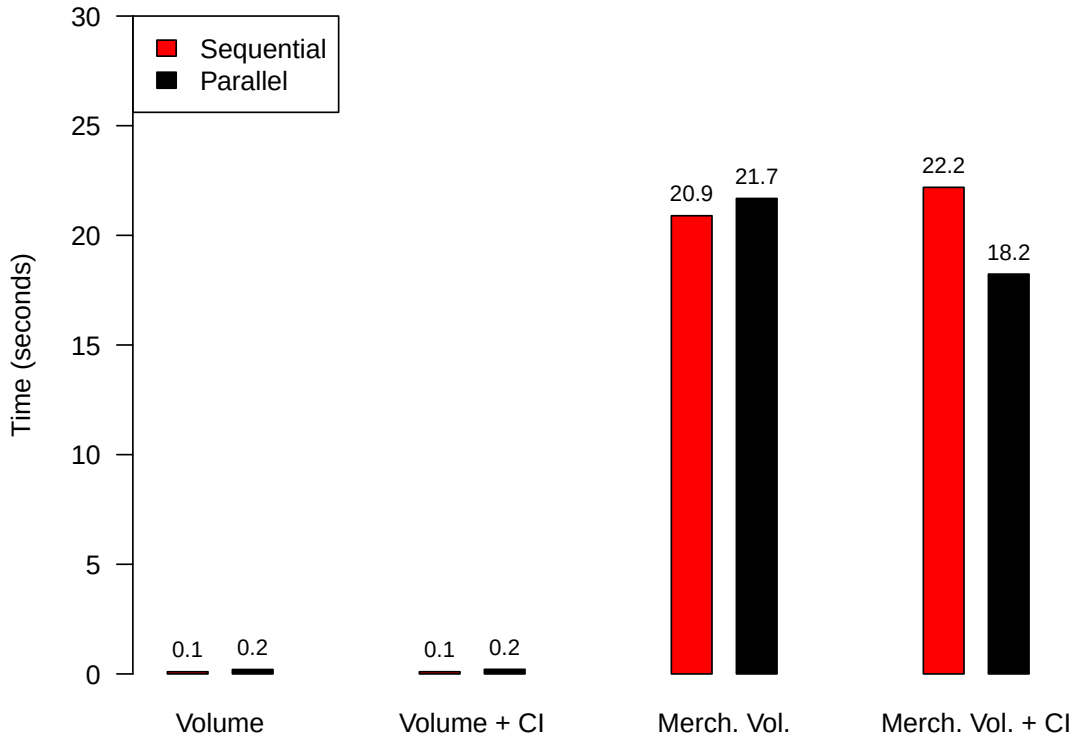

Figure 3.6: Comparisons of elapsed time for total and merchantable volume computation for a tree-list of size 10,000 using the package **PSTapeR**. Total volume and total volume plus confidence intervals on the left. Merchantable volume and merchantable volume plus confidence intervals on the right. The simulation was done using an iMac with processor Intel Core i3, 3.06GHz, and 8GB RAM.

3.6 Supplementary Material

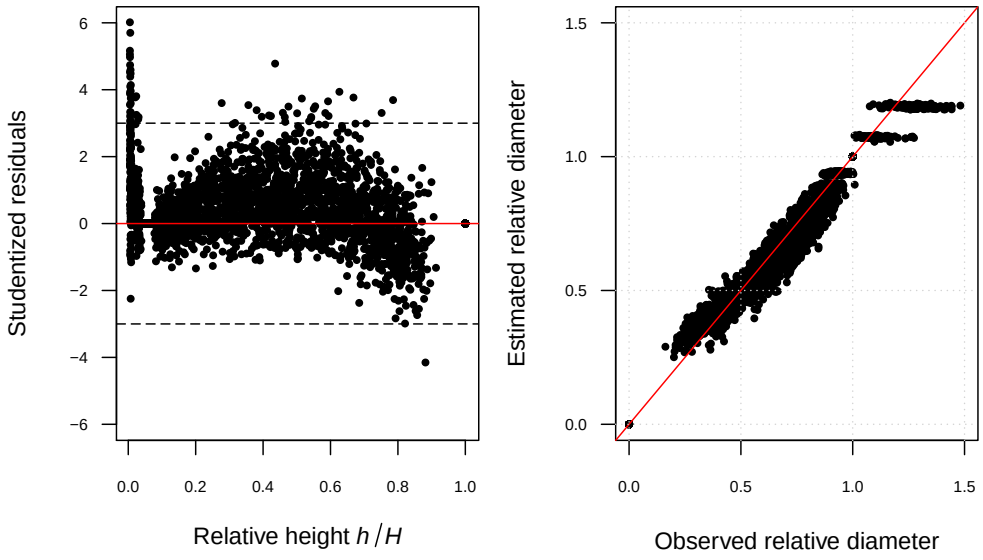

Figure 3.7: Studentized residuals (left), and comparison of predicted and observed relative diameters dob_{ij}/dbh_i (right) using the Sharma and Oderwald (2001) taper equation Eq. 3.14. Correction factors were used to obtain unbiased predictions of *dob* after retransformations of the predicted dependent variable as proposed by Czaplewski and Bruce (1990). The following segmented linear regression was used to predict the residual sample variance: $\hat{E}\left[\epsilon_{ij}^2\right] = 0.0003274(1-\frac{h_{ij}}{H_i})^2 - 0.0003662(1-\frac{h_{ij}}{H_i})^3 + 0.007137(1-\frac{h_{ij}}{H_i})I$; $I = 1$ if $(\frac{h_{ij}}{H_i}) < 0.0436$, 0 otherwise; 0.0436 is the join point fixed near the lower extreme of relative height for breast height measurements; $R^2 = 0.7041$; Residual standard error 0.00004218 with 53 degrees of freedom.

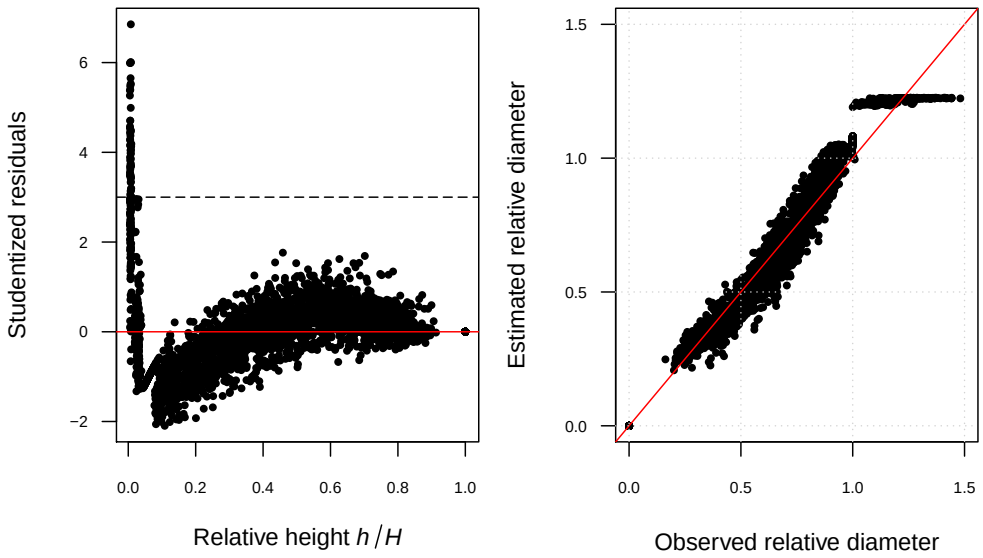

Figure 3.8: Studentized residuals (left), and comparison of predicted and observed relative diameters dob_{ij}/dbh_i (right) using the Kozak et al. (1969) taper equation Eq. 3.17. Correction factors were used to obtain unbiased predictions of *dob* after retransformations of the predicted dependent variable as proposed by Czaplewski and Bruce (1990). The following segmented linear regression was used to predict the residual sample variance: $\hat{E}\left[\epsilon_{ij}^2\right] = 0.009482(1-\frac{h_{ij}}{H_i}) - 0.007815(1-\frac{h_{ij}}{H_i})^3 + 1.008545(1-\frac{h_{ij}}{H_i})I$; $I = 1$ if $(\frac{h_{ij}}{H_i}) < 0.0436$, 0 otherwise; 0.0436 is the join point fixed near the lower extreme of relative height for breast height measurements; $R^2 = 0.9432$; Residual standard error 0.002698 with 53 degrees of freedom.

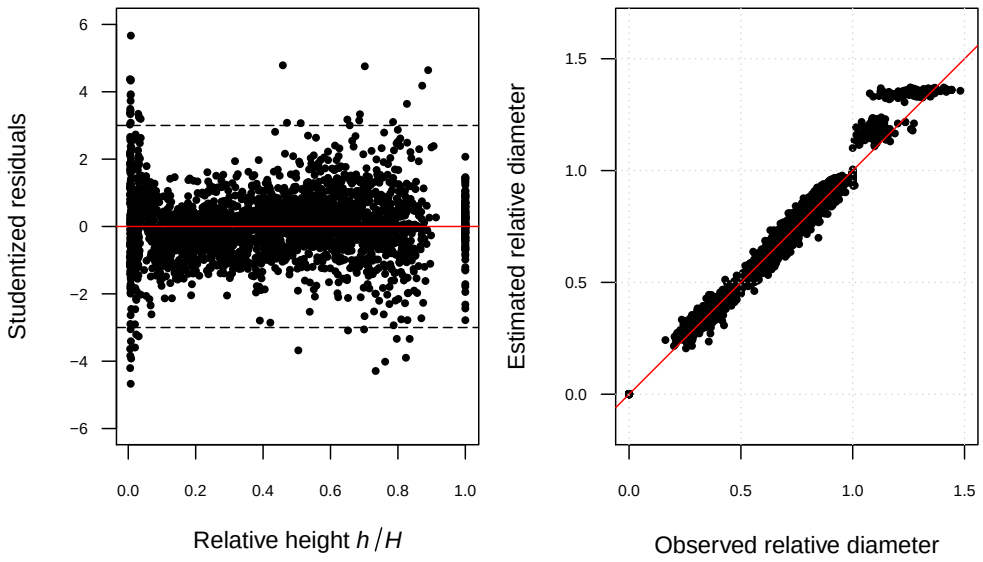

Figure 3.9: Studentized residuals (left), and comparison of predicted and observed relative diameters dob_{ij}/dbh_i (right) using the Max and Burkhart (1976) taper equation Eq. 3.19. Correction factors were used to obtain unbiased predictions of *dob* after retransformations of the predicted dependent variable as proposed by Czaplewski and Bruce (1990). The following segmented linear regression was used to predict the residual sample variance: $\hat{E}\left[\epsilon_{ij}^2\right] = 0.015159(1 - \frac{h_{ij}}{H_i})^2 - 0.014068(1 - \frac{h_{ij}}{H_i})^3 + 0.869980(1 - \frac{h_{ij}}{H_i})I$; $I = 1$ if $(\frac{h_{ij}}{H_i}) < 0.0436$, 0 otherwise; 0.0436 is the join point fixed near the lower extreme of relative height for breast height measurements; $R^2 = 0.9545$; Residual standard error 0.002 with 53 degrees of freedom.

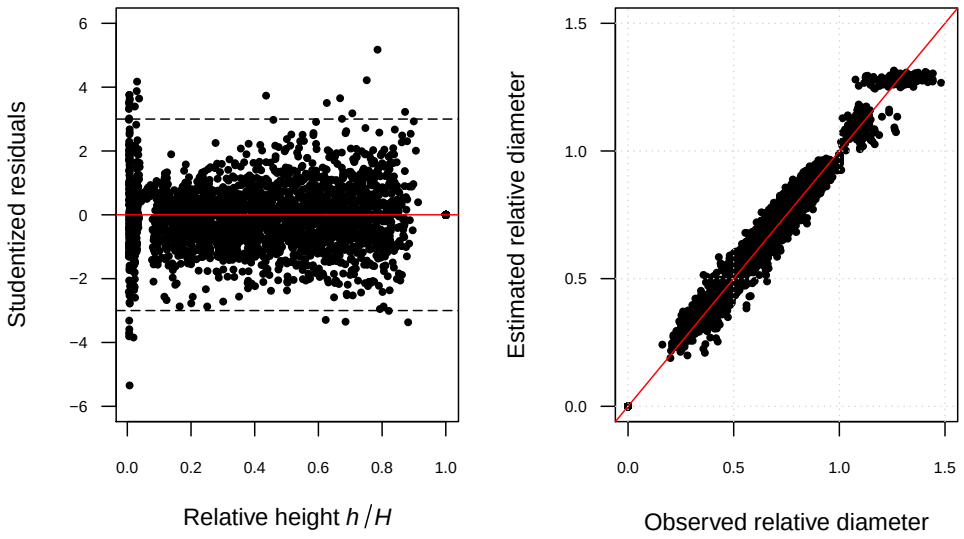

Figure 3.10: Studentized residuals (left), and comparison of predicted and observed relative diameters dob_{ij}/dbh_i (right) using the Fang et al. (2000) taper equation Eq. 3.21.

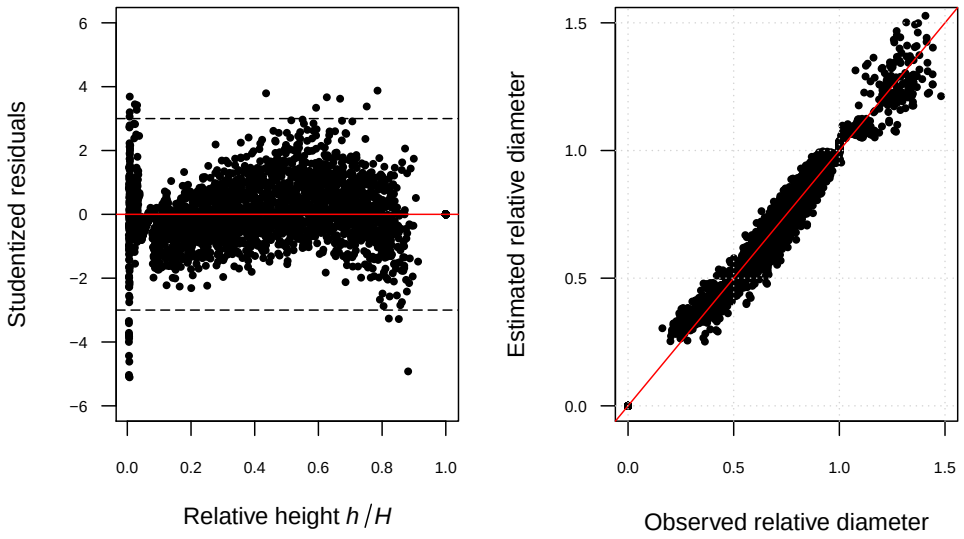

Figure 3.11: Studentized residuals (left), and comparison of predicted and observed relative diameters dob_{ij}/dbh_i (right) using the Kozak (2004) taper equation Eq. 3.23.

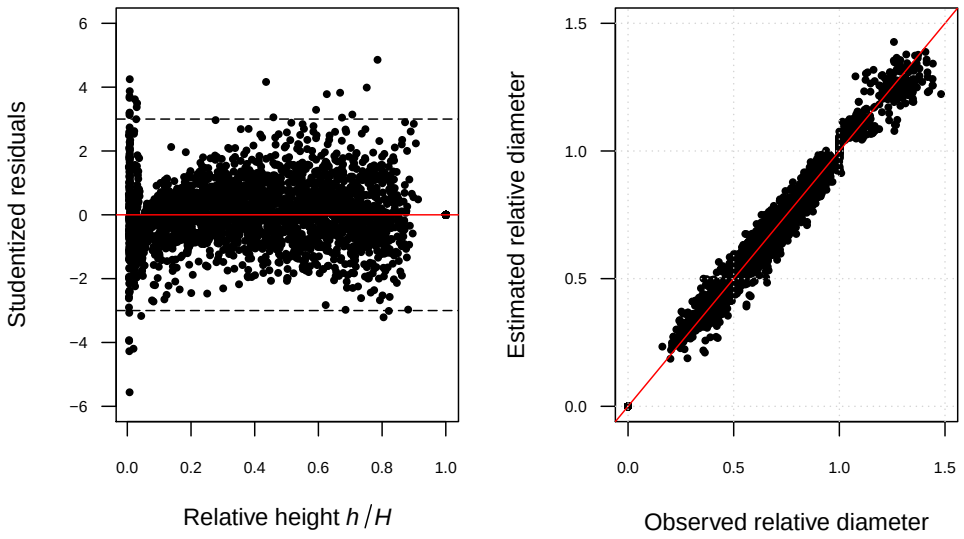

Figure 3.12: Studentized residuals (left), and comparison of predicted and observed relative diameters dob_{ij}/dbh_i (right) using the Kozak (2004) taper equation Eq. 3.24.

Table 3.8: Average bias (×1000) on volume outside bark predictions by relative height classes with six parametric taper equations and two P-Spline (Eq. 3.2 and Eq. 3.3). The lower bias in each relative height class is bolded.

Relative height	No. obs.	Model								
		Eq. 3.2	Eq. 3.3	Eq. 3.14	Eq. 3.17	Eq. 3.19	Eq. 3.19r	Eq. 3.21	Eq. 3.23	Eq. 3.24
		Mean Bias								
(0-0.1]	485	0.437	0.329	0.448	-0.506	0.959	0.988	0.838	**0.284**	-0.294
(0.1,0.2]	283	**0.016**	-0.195	-1.090	-8.253	-1.046	-0.962	-0.466	-2.406	-0.931
(0.2-0.3]	268	0.229	**0.038**	1.410	-5.001	-1.275	-1.161	-0.911	-1.223	0.537
(0.3-0.4]	268	0.587	0.407	2.739	-1.187	-0.479	**-0.352**	-0.424	0.481	0.951
(0.4-0.5]	268	0.789	0.674	3.288	1.729	0.429	0.554	**0.156**	1.925	1.086
(0.5-0.6]	271	0.896	0.725	3.154	3.564	1.125	1.252	**0.387**	2.733	1.146
(0.6-0.7]	268	0.634	0.547	1.948	3.721	1.134	1.249	**-0.022**	2.262	0.855
(0.7-0.8]	260	0.630	0.661	**0.171**	2.551	0.782	0.872	0.334	0.942	0.494
(0.8-0.9]	144	0.199	0.711	-1.510	1.005	0.571	0.647	0.313	-0.673	**0.177**
(0.9-1]	148	-3.313	**-0.861**	-7.029	-2.710	-2.154	-2.024	-2.331	-5.332	-2.737
All	2663	0.286	0.336	0.776	-0.528	0.146	0.242	**-0.058**	0.180	0.215

References

Arias-Rodil, M., Diéguez-Aranda, U., and Burkhart, H. (2017). Effects of measurement error in total tree height and upper-stem diameter on stem volume prediction. *Forest Science.*, 63(3):250–260.

Baskerville, G. (1972). Use of logarithmic regression in the estimation of plant biomass. *Canadian Journal of Forest Research.*, 2:49–53.

Burkhart, H., Avery, T., and Bullock, B. (2019). *Forest measurements*. Waveland Press, INC., Long Grove, Illinois, sixth edition edition.

Cao, Q. and Wang, J. (2011). Calibrating fixed- and mixed-effects taper equations. *Forest Ecology and Management.*, 262(4):671–673.

Czaplewski, R. and Bruce, D. (1990). Retransformation bias in a stem profile model. *Canadian Journal of Forest Research.*, 20:1623–1630.

de Boor, C. (2001). *A practical guide to splines*, volume 27 of *Applied Mathematical Science*. Springer New York. 346p.

Diéguez-Aranda, U., Castedo-Dorado, F., Álvarez González, J., and Rojo, A. (2006). Compatible taper function for scots pine plantations in northwestern spain. *Canadian Journal of Forest Research.*, 36(5):1190–1205.

Eilers, P. C. and Marx, B. (1996). Flexible smoothing with B -splines and penalties. *Statistical Science.*, 11(2):89–121.

Fang, Z. and Bailey, R. (1999). Compatible volume and taper models with coefficients for tropical species on hainan island in southern china. *Forest Science.*, 45(1):85–100.

Fang, Z., Borders, B., and Bailey, R. (2000). Compatible volume-taper models for loblolly and slash pine based on a system with segmented-stem form factors. *Forest Science.*, 46(1):1–12.

Figueiredo-Filho, A., Borders, B., and Hitch, K. (1996). Number of diameters required to represent stem profile using interpolated cubic splines. *Canadian Journal of Forest Research.*, 26:1113–1121.

Goulding, C. J. (1979). Cubic spline curves and calculation of volume of sectionally measured trees. *New Zealand Journal of Forestry Science.*, 9(1):89–99.

Gregoire, T., Schabenberger, O., and Kong, F. (2000). Prediction from an integrated regression equation: A forestry application. *Biometrics.*, 56(2):414–419.

Gu, C. (2013). *Smoothing spline ANOVA models*. Number 297 in 2nd ed. Springer series in statistics. Springer, New York. 433 p.

Hastie, T. (1996). Pseudosplines. *Journal of the Royal Statistical Society. Series B (Methodological).*, 58(2):379–396.

Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statistical Science.*, 1(3):297–318.

Haywood, J. (2005). Influence of precommercial thinning and fertilization on total stem volume and lower stem form of loblolly pine. *Southern Journal of Applied Forestry.*, 29(4):215–220.

Henningsen, A. and Hamann, J. (2007). Systemfit: a package for estimating systems of simultaneous equations in r. *Journal of Statistical Software.*, 23(4):1–40.

Huggett, R., Wear, D., Li, R., Coulston, J., and Liu, S. (2013). Forecasts of forest conditions. p. 73-101. In *The southern forest futures project: Technical report.*, number SRS-GTR-178. in Gen. Tech. Rep. USDA-Forest Service, Southern Research Station, Asheville, NC.

Koskela, L., Nummi, T., Wenzel, S., and Kivinen, V. (2006). On the analysis of cubic smoothing spline-based stem curve prediction for forest harvesters. *Canadian Journal of Forest Research.*, 36:2909–2919.

Kozak, A. (2004). My last words on taper equations. *The Forestry Chronicle.*, 80(4):507–515.

Kozak, A., Munro, D., and Smith, J. (1969). Taper functions and their application in forest inventory. *The Forestry Chronicle.*, 45(4):278–283.

Krivobokova, T., Crainiceanu, C., and Kauermann, G. (2008). Fast adaptive penalized splines. *Journal of Computational and Graphical Statistics.*, 17(1):1–20.

Krivobokova, T., Kneib, T., and Claeskens, G. (2010). Simultaneous confidence bands for penalized spline estimators. *Journal of the American Statistical Association.*, 105(490):852–863.

Kublin, E., Augustin, N., and Lappi, J. (2008). A flexible regression model for diameter prediction. *European Journal of Forest Research.*, 127:415–428.

Kublin, E., Breidenbach, J., and Kandler, G. (2013). A flexible stem taper and volume prediction method based on mixed-effects b-spline regression. *European Journal of Forest Research.*, 132:983–997.

Kuzelka, K. and Marusák, R. (2014). Use of nonparametric regression methods for developing a local stem form model. *Journal of Forest Science.*, 60(11):464–471.

Kuzelka, K. and Marusák, R. (2015). Input point distribution for regular stem form spline modeling. *Forest Systems.*, 24(1):1–8.

Lappi, J. (2006). A multivariate, nonparametric stem-curve prediction method. *Canadian Journal of Forest Research.*, 36(4):1017–1027.

Leites, L. and Robinson, A. (2004). Improving taper equations of loblolly pine with crown dimensions in a mixed-effect modeling framework. *Forest Science.*, 50(2):204–212.

Liu, C. (1980). Log volume estimation with spline approximation. *Forest Science.*, 26(3):361–369.

Max, T. and Burkhart, H. (1976). Segmented polynomial regression applied to taper equations. *Forest Science.*, 22(3):283–289.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and Team., R. C. (2018). nlme: Linear and nonlinear mixed effects models. R package version 3.1-137.

Robinson, A. P., Lane, S. E., and Thérien, G. (2011). Fitting forestry models using generalized additive models: a taper model example. *Canadian Journal of Forest Research.*, 41(10):1909–1916.

Robinson, K. (1991). That BLUP is a good thing: the estimation of random effects. *Statistical Science.*, 6(1):15–51.

Ruppert, D., Wand, M., and Carrol, R. (2003). *Semiparametric Regression.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, New York. 386 p.

Ruppert, D., Wand, M., and Carroll, R. (2009). Semiparametric regression during 2003 - 2007. *Electronic Journal of Statistics.*, 3:1193–1256.

Scolforo, H., McTague, J., Raimundo, M., Weiskittel, A., Carrero, O., and Scolforo, J. (2018). Comparison of taper functions applied to eucalypts of varying genetics in Brazil:

application and evaluation of the penalized mixed spline approach. *Canadian Journal of Forest Research.*, 48(5):568–580.

Sharma, M. and Oderwald, R. (2001). Dimensionally compatible volume and taper equations. *Canadian Journal of Forest Research.*, 31(5):797–803.

Sharma, M. and Zhang, S. (2004). Variable-exponent taper equations for jack pine, black spruce, and balsam fir in eastern Canada. *Forest Ecology and Management.*, 198:39–53.

Stage, A. and Crookston, N. (2007). Partitioning error components for accuracy-assessment of near-neighbor methods of imputation. *Forest Science.*, 53(1):62–72.

Trincado, G. and Burkhart, H. (2006). A generalized approach for modeling and localizing stem profile curves. *Forest Science.*, 52(6):670–682.

Wood, E., Bullock, B., Isik, F., and McKeand, S. (2015). Variation in stem taper and growth traits in a clonal trial of loblolly pine. *Forest Science.*, 61(1):76–82.

Chapter 4

Dynamic Stand Growth Model System for Loblolly Pine Responding to Mid-Rotation Treatments[1]

[1]Zapata-Cuartas, M., B.P. Bullock, C.R. Montes, and M. Kane (in preparation). Dynamic stand growth model system for loblolly pine responding to mid-rotation treatments. To be submitted to *Forests*.

Abstract

Intensive loblolly pine (*Pinus taeda* L.) plantation management in the southeastern United States includes mid-rotation silvicultural practices (MRSP) like thinning, fertilization, competitive vegetation control, and their combinations. Consistent and well-designed long-term studies considering interactions of MRSP are required to produce accurate projections and evaluate management decisions. Here we use longitudinal data from the regional Mid-Rotation Treatment study established by the Plantation Management Research Cooperative (PMRC) at the University of Georgia across the Southeast to fit and validate a new dynamic model system rooted in theoretical and biologically principles. A Weibull pdf was used as a modifier function coupled with the basal area growth model rather than a base-line yield model, resulting in a compatible growth and yield system. The model's parameters, standard deviation error functions, and cross-correlations were estimated simultaneously. Results indicated that the model projections reproduce the observed behavior of stand characteristics well and the extrapolations are plausible. Simulations across physiographic regions allow us to rank responses to combinations of treatments and compare them with the thinning only control. Because of the model structure, the response to treatment changes with location, age applied, and the dominant height growth as indicators of site quality. Therefore, the proposed model well represents regional growth conditions.

Keywords: Dynamic growth systems; fertilization; growth and yield modeling; non crop vegetation control; silvicultural treatment responses; thinning.

4.1 Introduction

Throughout the southeastern United States, pine plantations have made a significant expansion. Before 1952 the pine plantations in this region were marginal, but during 1952-2010, the planted area accumulated 15.8 million hectares (19 percent of the total forest area in

the Southeast) (Wear and Greis, 2012). Forecasts reveal a positive rate of change of conversion from natural regenerated pine types to pine plantations. From 2010 to 2060, pine plantations are expected to represent between 24 and 36 percent of the Southeastern forest area (Huggett et al., 2013). That means a net increment between 3 to 11.4 million hectares, mainly with loblolly pine (*Pinus taeda* L.) plantations. Loblolly pine is the most planted species in the southeastern U.S. and is the principal source of timber and forest products for the national forest industry.

Concurrent with this expansion, pine plantation management has increasingly turned more intensive. Precise improvements in silvicultural practices have shown remarkable productivity gain (Stanturf et al., 2003; Fox et al., 2004). Intensive management with Mid-Rotation Silvicultural Practices (MRSP) has been used extensively in the Southeast to raise the quality and proportion of volume in hight-value product classes in loblolly pine plantations. Here we will review typical MRSP like thinning (Thin), fertilization (Fert), and competitive vegetation control (CVC), and then we will explain how they have been incorporated in long-tern modeling.

Fertilization at mid-rotation with nitrogen and phosphorus has been well documented, and the typical response is a positive increasing productivity in loblolly pine plantations (Fox et al., 2007). Fert and CVC are common silviculture practices that serve to ameliorate nutrients and water limitations. Adding nutrients during the rotation produce an increase in nutrient availability (Fox et al., 2007) and has some effect on reducing the carbon cost to produce fine roots (Cannell and Dewar, 1994). MRSP with one-time application of nitrogen (N; 168 to 224 kilograms per hectare) and phosphorous (P; 28 to 56 kilograms per hectare) on most soil types throughout the southeastern U.S. produces a large and consistent growth response of stand volume of approximately 25 percent (Fox et al., 2007).

The response of loblolly pine to fertilization treatments can be explained by increasing nutrient availability at the tree level and, consequently, an increased leaf area index, resulting

in improved growth (Albaugh et al., 2017). Therefore, the magnitude of fertilizer response is associated with the gap between the leaf area index before the treatment application and the potential leaf area index that an operational plantation can develop (around LAI=4) (Sampson and Allen, 1999; Fox et al., 2007). N and P fertilization with responses of approximately 3.85 m^3 ha^{-1} yr^{-1} have been shown to enhance the profitability of loblolly pine plantations (Fox et al., 2007).

At early-ages herbaceous and hardwood competition can adversely affect growth (Jokela et al., 2010). Early control of competing woody and herbaceous vegetation increases wood volume by 23-121 percent when evaluated at age 15 (Miller et al., 2003). Although there is considerable literature examining CVC's effect at stand establishment, few works quantify the beneficial effects of CVC applied at the mid-rotation ages. The treatment consists of removing either the non-crop woody vegetation, the herbaceous weed vegetation, or both. The principle is to make the scarce resources available to the crop trees. Specifically, at the tree level, competing vegetation reduces nutrient availability and increases water competition, influencing radiation use efficiency.

Thinning is another important MRSP in loblolly pine plantation management. Thinning not only changes the structure of the stand (diameter distributions), but alters the post-thinning height growth of dominant and co-dominant trees (Sharma et al., 2006), reduces competition, and improve the stem quality of the residual stand trees. Thinning is a way to modify the process of site resource allocation. In general, the principal objective of this treatment is to concentrate light, water, and nutrients on fewer and better trees that can survive over the entire rotation, ending up with higher valued stems.

Foresters and landowners are particularly interested in knowing the long-term response of important stand characteristics after MRSP. Therefore, existing empirical growth and yield models have been adapted to include treatment indicators, timing, and intensity parameters to achieve the desired adjusted predictions. For the case of MRSP analyzed here, most

of the empirical approaches reviewed invariably underpin the following two assumptions: First, an adequate yield and growth model is available to simulate the untreated condition. Second, the yield after MRSP can be expressed as a function of the estimated untreated yield or yield-baseline (also known as reference curve) and a multiplicative or additive modifier sub-model (Weiskittel, 2011). Thus, after treatment, prediction accuracy depends on how well the yield-baseline perform for a particular locality or species and how well the modifier undertakes the assumptions on the type and size of response regarding treatment(s).

In many publications concerning stand response to MRSP, it is standard to use multiplicative or additive sub-models, also called "modifiers", to force a base-line model to predict the expected growth after treatment (Pienaar, 1997; Amateis, 2000; Gyawali and Burkhart, 2015). Therefore, it is assumed that the modifier-submodel corrects the projections attained with the yield-baseline to reflect the new growth condition or resource availability.

There are different properties and assumptions used with those modifiers. Table 4.1 summarizes the most important characteristics of some modifiers used in the literature. For example, the treatment's duration effect Z (lasting effect of the treatment in years since treatment), or if the sub-model parameters are estimated simultaneously with the base-line model or are estimated independently of the base-line model.

While the additive and multiplicative modifiers have shown to be a practical solution to modeling MRSP, they have the drawback that the resulting yield model is not compatible with the growth baseline model. The concept of compatible growth and yield model was introduced by Clutter (1963). This principle consists of that growth and yield are not two separate attributes but are closely related to one another regarding that yield is obtained by mathematically integrating the growth equation over time.

The reviewed literature so far allows us to highlight two research gaps in growth and yield modeling. First, the lack of prediction and projection models for loblolly pine derived from consistent and well-designed long-term studies considering interactions of MRSP. Second,

the modifiers used so far are coupled with a base-line yield model rather than a base-line growth model, resulting in a non-compatible growth and yield expression. Our premise is that it makes more biological sense to consider the MRSP response as a modifier of the stand growth. The yield then results from the integration of the projected growth after treatment.

In an effort to increase understandings of the intensive management with MRSP in terms of growth, product distributions, and financial returns in the Southern loblolly pine plantations, the Plantation Management Research Cooperative (PMRC) at the University of Georgia, established a regional Mid-Rotation Treatment Study (MRT) with a set of 49 experimental locations in non-thinned and first-thinned loblolly pine plantations across the southeastern U.S. between 2009-2018. The study's goals were to (1) Develop databases appropriate for modeling first- and second-thinned plantations; (2) Develop response models for fertilization and competitive vegetation control treatments of first- and second-thinned plantations. This present research address goal (2) of the study.

Alternative and flexible growth modeling strategies need to be addressed to overcome the second gap mentioned above. This study proposes a dynamic state-space approximation. Dynamic growth and yield systems use the current state of the stand attributes and predict the rate of change in the state using first-order differential equations. The state-space approach has proven to be robust and performed very well for loblolly pine (Garcia et al., 2011).

The main goals in this research were to 1) propose a new whole-stand dynamic growth and yield system for thinned and non-thinned loblolly pine plantations rooted in theoretical and biologically consistent state-space modeling, 2) to extend the dynamic system with a robust growth response modifier for Fert and CVC treatments, singly and in combination, following first- and second-thinning operations. The expectation with these new systems of models is to provide accurate stand predictions and projections to reflect silvicultural

treatments impacts and better support silvicultural investment and forest planning decision making.

4.2 Materials and Methods

4.2.1 Study Design

At any given location, the MRT study design was a randomized 2×2 factorial of post-thinning Fert and CVC plus one control plot, with no replication within a location.

The study covers two southern physiographic regions; the Upper Coastal Plain/Piedmont (UC/PI) and the Lower Coastal Plain (LC). Within each physiographic group, plots were established in first-thinned and second-thinned loblolly pine plantations. Within each thinning condition and region combination, three stands (Installations) were identified for each of four unique combinations of site index at a base age of 25 years (low: 16.8-21.3 m and high: 21.3-30.5 m) and initial stand basal area (low: 20.7-27.6 m^2 ha^{-1} and high: 27.6-34.4 m^2 ha^{-1}), resulting in 12 stands per thinning condition and region combination. One of the three stands for a given site index and initial basal area class was assigned to a post-thinning basal area of 11.5 m^2 ha^{-1}, a second stand to a post-thinning basal area of 16 m^2 ha^{-1}, and the third stand to a post-thinning basal area of 20.7 m^2 ha^{-1}.

Within each stand (installation), five plots were established. One plot did not receive any subsequent thinning or cultural treatment, and four plots were thinned to the basal area target. After thinning, the following treatments were randomly assigned to the four remaining plots; no treatment, competitive vegetation control (CVC) only, fertilization only (Fert), and CVC plus Fert (see Table 4.2). The same treatment levels were applied to all the installations.

The complete experiment was designed to have 48 installations and a total of 60 plots in each combination of the thinning condition and region (2 site index classes x 2 initial

Table 4.1: Some stand-level modifiers used to predict Mid-Rotation Silvicultural Practices (MRSP) responses in growth and yield models reported in the southeastern U.S. Thin = thinning, Fert = fertilization, BA = basal area, N = trees per hectare, MV = merchantable volume, M = multiplicative modifier, A = additive modifier, Z = lasting effect of the treatment in years since treatment, k = fixed maximum lasting time effect.

MRSP	**Source**	**Stand atribute**	**Equation**	**Type of modifier**	**Base-line Yield model**	**Duration effect**	**Estimation Procedure Base-line and Modifier**
Thin	Pienaar (1997)	BA	eq. (4.55)	M	Counterpart stand	$0< Z <$inf	Independently
Thin	Amateis (2000)	BA	eq. (4.56)	A	Non-thinned stand	$0 <Z <$inf	Independently
Thin		N	eq. (4.57)	A	Non-thinned stand	$0 <Z <$inf	Independently
Thin	Gyawali and Burkhart (2015)	BA	eq. (4.58)	M	Non-thinned stand	$0 <Z <k$	Simultaneously
Thin		N	eq. (4.59)	M	Non-thinned stand	$0 <Z <k$	Simultaneously
Fert	Ballard (1984)	MV	eq. (4.60)	M	Non-thinned stand	$0 <Z <$inf	Independently
Fert	Gyawali and Burkhart (2015)	BA	eq. (4.61)	M	Non-thinned stand	$0 <Z <$inf	Simultaneously
Fert	Scolforo et al. (2020)	V	eq. (4.62)	M	Non-thinned stand	$0 <Z <$inf	independently
Fert	Amateis et al. (2000)	DH	eq. (4.63)	A	Non-thinned stand	$0 <Z <k$	independently
Fert	Amateis et al. (2000)	BA	eq. (4.64)	A	Non-thinned stand	$0 <Z <k$	independently

Table 4.2: Treatments description for the MRT study.

Treatment number	Thinning	Silvicultural treatment
1 - Control	No	None
2 - Thin	Yes	None
3 - Fert	Yes	Fertilize (224.2 kg N ha^{-1} + 28 kg P ha^{-1})
4 - CVC	Yes	Prescribed for each location as conditions warrant and with follow-up to obtain excellent operational efficacy
5 - Fert + CVC	Yes	Fertilize + CVC

basal area classes x 3 post-thinning basal areas x 5 treatments). One plot represents each unique combination of site index, pre-thinning basal area, and thinning and post-thinning treatment regime. The plot size for first-thinning and second-thinning installations was 3000 m^2 gross plot with a 2000 m^2 measurement plot. This allowed for approximately 50 trees per measurement plot for the most intensive second thinning option and more than 50 trees per measurement plot for the other thinning options. The MRT experiment dataset currently consist of 49 installations. Table 4.3 shows the distribution of plots by physiographic region for these 49 installations.

Figure 4.1 shows the spatial distribution of 49 installations from the MRT study across the southeastern United States.

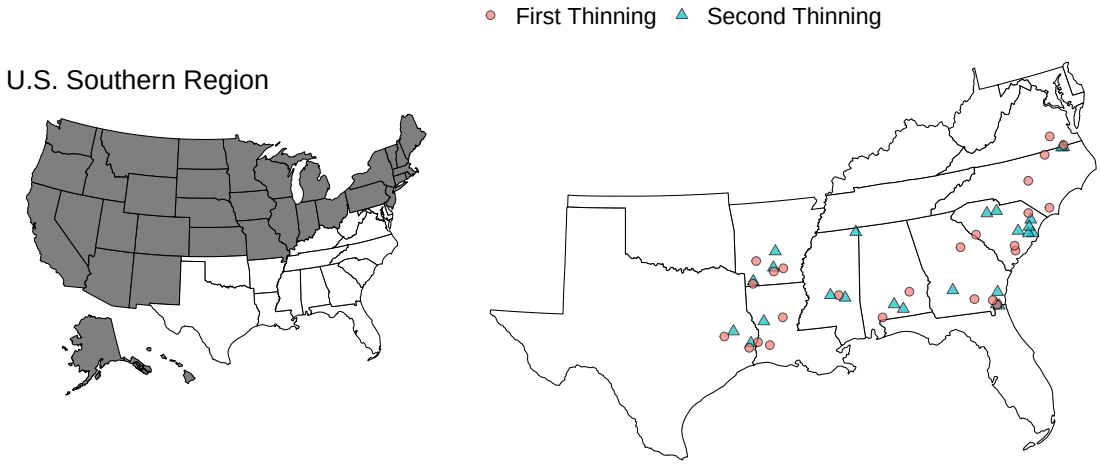

Figure 4.1: The southeastern U.S. region (left) and location of 49 installations from the PMRC Mid-Rotation Treatment study (right). Twenty-five installations in first-thinned and 24 in second-thinned loblolly pine plantations. In each installation, an unreplicated randomized 2×2 factorial setting for post-thinning fertilization and competitive vegetation control treatments was established plus a control non-thinned plot.

The first-thinning was done with an improvement cut from below (fifth row removed with a selection from below in leave rows), removing mostly smaller trees with some large trees taken for spacing or low stem quality. The second-thinning method was by selection from below while considering spatial distribution. Other information related to tree quality and presence of defects or diseases was also collected.

All plots were measured before thinning, immediately post-thinning (as needed), and every two years after that. Diameters at breast height (DBH) were measured at 1.37 m to the nearest 0.25 cm. The total height (H) measurements were collected using Haglöf Vertex

hypsometers to the nearest 0.3 m on a subsample of trees selected from the tree diameter distribution.

Plots that received the first-thinning treatment were established in plantations no more older than 17 years, and the locations were selected across the regions in a such way that the range of the stand attributes were similar. The age and intensity of the first-thinning in the selected plots for second thinning was unknown but the current stand conditions were used in site selection. The second-thinning installations have a larger age range at establishment (16 to 32 years). Table 4.3 shows the MRTs summary and ranges for age (Age), dominant height (DH), stand density (N), and basal area (BA) by physiographic region and thinning condition at the plot establishment. In general the selected 5 plots within an installation were homogeneous prior to treatment.

Table 4.3: Number of plots and range values of the MRT at establishment. Age in years, dominant height (DH) in m, stand density (N) in trees ha^{-1}, and basal area (BA) in m^2 ha^{-1}.

	First-thinning					Second-thinning				
Region	**No. Plots**	**Age**	DH	N	BA	**No. Plots**	**Age**	DH	N	BA
LC	60	11-17	14.6-21.2	1013-1735	10.8-44.7	55	16-25	17.6-25.6	430-717	11.4-33.7
PI	20	14-15	12.9-21.1	1310-1626	11.1-44.6	10	21-27	19.6-23.1	534-578	15.6-33.2
UC	45	12-17	13.8-19.6	919-1670	9.9-45.5	55	16-32	17.3-31.3	247-776	10.7-36.9

The CVC treatment was prescribed for each location as conditions warrant. This treatment removed all non-pine competitive vegetation after thinning and with one follow-up for missed areas to enhance application uniformity. The fertilization treatment was timed to coincide with or follow the CVC treatment to maximize the fertilization effect. Standard doses of nitrogen (N) and phosphorous (P) per hectare were applied to all those plots receiving the fertilizer treatment.

Before the CVC treatment, this study measured the competition in subplots covering approximately 5% of the measurement plot. The subsample measurements included large and small arborescent, shrubs count, and weeds groundcover percent. Competition vegetation

measurements were not used in the development of the stand level growth models proposed here.

4.2.2 Model Description

The dynamic model system proposed here is rooted in theoretical and biologically consistent growth principles. We have used available knowledge from previous published empirical and dynamic models to formulate a new system capable of simulating stand state attribute behavior and growth responses to MRSP. The model uses a state-space approach where each state is represented as an ordinary differential equation. The system includes dominant height (DH), stand density as trees per hectare (N), and stand basal area (BA). We used the MRT study dataset to estimate system parameters. We used diagnostic graphics and statistics tests to select experiment factor variables for each state model independently, and then, the ensemble model was fitted simultaneously. Next, we describe the proposed model for each state and the system parameter fitting method.

Dominant Height (DH) model

The DH is the state-driven component for N and BA, following the same argumentation presented by Garcia et al. (2011). We define DH as the average tree height for those trees which DBH is larger than the plot average DBH. However, when thinned, the average DBH is altered, producing thinning-dependent DH measurements. For this reason, we calculated the average of tree heights for those trees which DBH is larger than the average DBH and remain in the stand after thinning, and that more likely will survive at the rotation age; that is, those trees without annotations of damages or diseases after thinning. The base-line growth model form for DH follows the von Bertalanffy differential equation (Bertalanffy, 1949),

$$\frac{dDH}{dt} = \alpha DH^m - kDH \tag{4.1}$$

where α, m, and k are parameters, and t is the independent time variable in years. This model states that the change in DH over time is the result of confronting the anabolic process, αDH^m that makes the trees grow in height, and a catabolic process kDH, that limits potential height growth. The anabolic process is proportional to an allometric relationship with the stand DH. The catabolic process is assumed proportional to the stand DH size. This model was introduced in the forestry literature by Pienaar and Turnbull (1973). The integral form or model in (4.1) has been used to model the DH yield (Garcia, 1983) and the basal area (Pienaar and Turnbull, 1973).

Further, we found that the parameters α and m can be function of experimental conditions. The following expressions were used,

$$\boldsymbol{\alpha} = a_0 + a_1 R + a_2 IUC + a_3 IPI \tag{4.2}$$

$$\mathbf{m} = a_4 + a_5 R + a_6 IUC + a_7 IPI \tag{4.3}$$

where $a_0, a_1, \ldots, a_7$ are parameters, R is the thinning intensity defined as the ratio of trees per hectare removed at the moment of the thinning and the trees per hectare before thinning, IUC is an indicator variable that takes the value of 1 if the plot belongs from the UC region and it takes the value of zero otherwise, IPI is another indicator variable that takes the value of 1 if the plot belongs from the PI regions or takes a zero otherwise, and LC is the contrast factor level. Note that for the control plots, the amount of trees per hectare removed is zero, and then, the effect of R in (4.2) and (4.3) is goes away. The complete differential form for DH is

$$\frac{dDH}{dt} = \boldsymbol{\alpha} DH^{\mathbf{m}} - kDH \tag{4.4}$$

Equation (4.4) is a nonlinear ordinary differential equation. A projection equation for DH can be obtained from the integration of (4.4) applying Bernulli's integration equation (Blanchard et al., 2012) (see Appendix 4.6.1),

$$DH_2 = \left[\frac{\boldsymbol{\alpha}}{k} + \left[DH_{th}^{(1-\mathbf{m})} - \frac{\boldsymbol{\alpha}}{k}\right] exp\{-k(1-\mathbf{m})(t-t_{th})\}\right]^{\frac{1}{1-\mathbf{m}}} \tag{4.5}$$

where DH_{th} is the known dominant height determined before thinning when $t = t_{th}$. The prediction equation can be derived using the origin solution $(0, 0)$ for (t_{th}, DH_{th}). Fixing the base age at 25 years into (4.5) results in a Site Index (SI) equation and solving for DH results an equation in terms of Site Indexprediction equation in terms of SI (see Appendix 4.6.1),

$$DH = \left(\frac{\boldsymbol{\alpha}}{k}\right)^{\frac{1}{(1-\mathbf{m})}} \left[1 - \left(1 - SI^{(1-\mathbf{m})}\frac{k}{\boldsymbol{\alpha}}\right)^{\frac{t}{25}}\right]^{\frac{1}{(1-\mathbf{m})}} \tag{4.6}$$

Stand density

Stand density can change by natural density-dependent mortality or by intervention with thinning. It is thought that the course and magnitude of change in stand density must be affected by the intensity of thinning. We adopted a flexible model proposed by (Garcia et al., 2011) and modified parameters as a function of experimental factors and thinning intensity,

$$\frac{dN}{dDH} = -\left(\frac{n_1}{1e+16}\right) DH^{n_2} N^{\mathbf{n_3}} \tag{4.7}$$

$$\mathbf{n_3} = n_{30} + n_{31}R \tag{4.8}$$

where N is number of trees per hectare, n_1, n_2, n_{30}, and n_{31} are parameters. Here we are assuming that the original parameter n_3 now changes linearly with the thinning intensity to reflect reduced mortality typically observed after thinning. This model form is similar to the model proposed by Clutter and Jones (1980), but using DH as the independent variable. The selection of DH instead of the chronological age as a determinant of stand density changes has several benefits toward understanding the mortality. On sites with high site quality and fast growth rates it is expected that there is an early occurrence of intraspecific competition accelerating the density-dependent mortality process. DH growth rather than the lineal time increment may better express the causal effect of stand size mortality (Garcia et al., 2011). Equation (4.7) then states that the change in trees per hectare is proportional to a DH allometric term and with the instantaneous stand density. Note that the power term for N in (4.7) includes the thinning intensity R; this allows the model to express different future density (trees per hectare) effects on mortality for stands that having the same level of occupancy, come from different density management regimenes.

Integration of (4.7) gives an expression to project the stand density N_2 at the time when the stand has a DH_2, and given a known initial condition (DH_1, N_1). The projection form is (see Appendix 4.6.2)

$$N_2 = \left[N_1^{1-\mathbf{n_3}} + \left(\frac{n_1}{1e+16} \right) \frac{\mathbf{n_3} - 1}{n_2 + 1} \left[DH_2^{n_2+1} - DH_1^{n_2+1} \right] \right]^{\frac{1}{(1-\mathbf{n_3})}} \tag{4.9}$$

Basal area

We consider the DH growth as a driver for modeling the change in basal area. Inspired in the work presented by Garcia et al. (2011), we hypothesized that the change in the basal area results from comparing the potential total gross increment and a detrimental component. The gross increment is proportional to the product of two allometric expressions:

$$\text{gross increment} = b_0 N^{b_1} BA^{b_2} \tag{4.10}$$

where b_0, b_1, and b_2 are parameters. We state that the gross increment depends on the instantaneous accumulation of stem-wood (a size effect measured as BA) and an allometric form for the available living trees in the stand.

Further, we consider that the allometric parameter b_1 can be extended as a linear function, including the effect of previous thinning managements and the actual thinning intensity.

The detriment component is assumed to be proportional to the amount of basal area from the living trees that is reduced relative to the DH growth, that is:

$$\text{detriment in } BA = b_3 \frac{BA}{N} \frac{dN}{dDH} \tag{4.11}$$

where b_3 is the proportional parameter, which also can be interpreted as the mean size of dying trees relative to the mean size of living trees. It is expected that for a growth period (e.g., two years), the total gross increment surpasses any losse due to mortality. Previous exploratory data analysis (not-shown here) indicated that thinning practices considerably reduce the mortality rate but favor the average tree basal area increment. Therefore the proportional parameter in (4.11) should reflect the stand management condition. The experimental factors in the MRT study were thoroughly evaluated through regression methods and graphical diagnostics to select the variables to be included in the expressions for b_1 and b_3, but that detail is omitted here. The final expression for b_3 is

$$\mathbf{b_3} = b_{30} + b_{31} IT + b_{32} R \tag{4.12}$$

where b_{30}, b_{31}, and b_{32} are parameters, IT is an indicator variable for the thinning condition, IT takes a value of 1 if the stand receives a second-thinning, and takes the value of

zero if it receives the first-thinning. For convenience (see Appendix 4.6.3) parameter b_1 is expressed as $b_1 = \mathbf{b_3}(b_2 - 1)$.

We hypothesized that BA growth might respond to the MRSP differently according to the physiographic region location, the plantation age when the treatment was applied, and the MRSP must have a temporal effect. To include this behavior in (4.10), the proportional parameter b_0 was extended as a linear combination of the physiographic region and a modifier equation. This modifier is a function of the type of MRSP present and the age when applied. The expression for b_0 is

$$\begin{aligned} \mathbf{b_0} &= b_{01} + b_{02}IUC + b_{03}IPI + \\ &\quad \frac{b_{04}}{Age_{th}}(If + Ir + Ifr) \times \\ &\quad \left[\left(\frac{\mathbf{c}}{\mathbf{d}}\right)\left(\frac{DH - DH_{th}}{\mathbf{d}}\right)^{\mathbf{c}-1} exp\left\{-\left(\frac{DH - DH_{th}}{\mathbf{d}}\right)^{\mathbf{c}}\right\}\right] \end{aligned} \tag{4.13}$$

$$\mathbf{c} = 1 + c_1 If + c_2 Ir + c_3 Ifr \tag{4.14}$$

$$\mathbf{d} = 1 + d_1 If + d_2 Ir + d_3 Ifr \tag{4.15}$$

where $b_{01}, \ldots, b_{04}$, and c_1, c_2, c_3, d_1, d_2, and d_3 are parameters, Age_{th} is the plantation age when the MRSP was applied. If, Ir, and Ifr are indicator variable for Fert, CVC, or Fert+CVC, respectively. These indicator variables take the value of one if the respective MRSP is present after thinning and take a value of zero elsewhere. The other variables as defined before.

The equation in (4.13) uses a modification of a two-parameter Weibull probability density function (pdf) to define a modifier function for each of the treatments like Fert, CVC, or their

combination. Note that this modifier is coupled with a BA growth model parameter instead of the BA yield. Here, the modifier is considered an intervention function that produces temporal growth increments compared to the thinning only control.

The Weibull pdf has been used before as a yield modifier equation (Gyawali and Burkhart, 2015). The Weibull pdf is flexible and can take multiples densities form, which here is used to represent different types of treatment responses on the proportionality parameter $\mathbf{b_0}$. Specifically, the scale parameter $\mathbf{d}$ is related to the lasting effect of the treatment in the thinned stand, and the shape parameter $\mathbf{c}$ is related to the mechanism by which the treatment interacts with the growth process. For example, unimodal pdf shapes indicate that the treatment's resource availability does not produce an immediate effect on growth but gradually accelerated until a maximum in response is achieved with a posterior descending as these extra resources are draining or used by the stand. Alternatively, an inverse j-shape type indicates an immediate growth response with a subsequent decreasing effect. When the pdf takes values close to zero, that implies that the stand growth follows the same growth rate of a thinning only control.

The Weibull pdf proposed here is a new conceptualization in this approach. Note that we used dominant height instead of the time elapsed from when the treatment was applied. There are two primary reasons for that. First, the use of $DH - DH_{th}$ in the Weibull modifier is consistent because DH is the independent variable in the BA growth model. Second, the resources provided by the treatments are used for growth, and the growth response behavior is guided by how the stand DH responds instead of following a chronological time-line dimension, which does not necessarily hold a strong causal relationships with the stand development.

The growth model for BA considered here was, therefore, of the form

$$\frac{dBA}{dDH} = \mathbf{b_0} N^{\mathbf{b_3}(b_2-1)} BA^{b_2} - \mathbf{b_3} \frac{BA}{N} \frac{dN}{dDH} \tag{4.16}$$

The integral of the differential equation (4.16) has a closed-form solution (see Appendix 4.6.3),

$$BA_2 = \Bigg[\quad BA_1^{(1-b_2)} \left(\frac{N_{th}}{N_2}\right)^{\mathbf{b_3}(1-b_2)} + (b_{01} + b_{02}IUC + b_{03}IPI)(1-b_2)N_2^{\mathbf{b_3}(b_2-1)}(DH_2 - DH_{th}) + \frac{b_{04}}{Age_{th}}(If + Ir + Ifr)\left(1 - exp\left\{-\left(\frac{DH_2 - DH_{th}}{\mathbf{d}}\right)^{\mathbf{c}}\right\}\right)(1-b_2)N_2^{\mathbf{b_3}(b_2-1)} \Bigg]^{\frac{1}{(1-b_2)}} \quad (4.17)$$

where BA_2 is the projected basal area when the stand has a stand density N_2 and a dominant height DH_2. We assume that the stand information immediately after thinning is known, that is, N_{th}, DH_{th}, and Age_{th} are known. Other variables and parameters are as defined previously.

4.2.3 Estimation

The model parameters for the three differential equations (4.1), (4.7), and (4.16) were initially fitted separately to obtain good initial values and as initial testing of the working hypothesis. AIC (Akaike's information criterion) and BIC (Schwartz's Bayesian information criterion) were used to decide on the inclusion of experimental factors or covariates in each model. Given the results of the initial separate fitting, then the system of growth equations was simultaneously estimated. All the calculations were performed in the statistical software language R (R Core Team, 2020) with the function `optim` and optimization algorithm `BFGS` for separate models and the algorithm `L-BFGS-B` for the complete system. The parameter estimation method proceeded as follow:

1. Obtain initial parameter estimations for the DH growth model independently of the other state variables, assuming the resulting yield follows a normal distribution with

constant variance and mean explained by equation (4.5). The initial value condition for each plot corresponds with the measured attributes at the thinning age.

2. Project the plot DH yield and use these projections (instrumental variables) for the initial parameter estimations for finding the stand density growth equation (4.7).

3. Use the yield projections of DH and N from the two previous steps and estimate parameters for the basal area growth model (4.1).

4. Use the plot prediction residuals to explore and propose a function that predicts the standard deviations model error for each state.

5. Use the previous plot level yield predictions to explore the type of cross-equation correlation for each combination of two responses.

6. Use the multivariate normal distribution framework. The projected states are the mean, and the estimated standard deviations and cross-equation correlation are used to construct the respective covariance matrix for each plot observation.

7. Use the maximum likelihood procedure to optimize the system parameters.

8. The inverse of the Hessian matrix resulting from the optimization process should be used to obtain approximate standard errors for parameters. We found that the resulting Hessian matrix is invertible but, some of the diagonal elements produce negative values and a few returned huge numbers. In this case, the likelihood function may still contain considerable information about the models of interest, but through the Hessian the parameters are non-identifiable Kreutz et al. (2013). As not all the parameters were completely identifiable with the inverse of the Hessian matrix, we then compute the standard errors using a boostrap method. Five hundred random replications of the complete experiment data set were obtained sampling with replacement within physiographic region and thinning condition, then the model system was fitted with

each replication and the fitted parameters were stored. The approximated standard error was obtained as:

$$SE_{Boot}(\theta_j) = \sqrt{\frac{\sum_{i=1}^{N}(\theta_{ij}^{*} - \hat{\theta}_j)^2}{N}} \tag{4.18}$$

where, $SE_{Boot}(\theta_j)$ is the bootstrap standard error for the jth parameter, θ_{ij}^{*} is the estimate jth paramater from the ith replicate data set, $\hat{\theta}_j$ is the estimate parameter with the original data set, N is the number of replications.

We defined the following functions to describe the standard deviations of state projections.

$$sd(DH_2) = \left[exp(\gamma_1)DH(t_2 - Age_{th})^{exp(\gamma_2)}\right]^{\gamma_3} \tag{4.19}$$

$$sd(N_2) = \gamma_4 + \gamma_5(DH_2 - DH_{th}) + \gamma_6(DH_2 - DH_{th})IT_1 \tag{4.20}$$

$$sd(BA_2) = \gamma_7 + \gamma_8(DH_2 - DH_{th}) + \gamma_9 BA_2 IT_2 \tag{4.21}$$

where $\gamma_1, \ldots, \gamma_9$ are parameters. The model in (4.19) assumes the prediction standard deviation is proportional to the response and increases as the projection length increases. The increased projection length is scaled to determine if the standard deviation increases in relation to projection length in a linear ($\gamma_2 = 0$), exponential ($\gamma_2 > 0$), or logarithmic ($\gamma_2 < 0$) fashion (Gallagher, 2019). The model in (4.20) assumes the prediction standard deviation is lineal with the increment in dominant height since thinning and with different slopes for the control ($IT_1 = 1$). The model in (4.21) assumes the prediction standard deviation is lineal with the increment in dominant height since thinning and with an additive effect proportional to the second thinning plots ($IT_2 = 1$).

In this study, the cross-equation correlation was found simultaneously with the mean functions and the standard deviations for each of the three response variables. The correlation matrix for the three state variables was defined as follow:

$$\mathbf{\Phi} = \begin{bmatrix} 1 & \phi_{DH\_N} & \phi_{DH\_BA} \\ \phi_{DH\_N} & 1 & \phi_{N\_BA} \\ \phi_{DH\_BA} & \phi_{N\_BA} & 1 \end{bmatrix} \tag{4.22}$$

The standard deviations of each state were expressed as a diagonal matrix as follow:

$$\mathbf{s} = \begin{bmatrix} sd(DH_t) & 0 & 0 \\ 0 & sd(N_t) & 0 \\ 0 & 0 & sd(BA_t) \end{bmatrix} \tag{4.23}$$

The covariance matrix used in the likelihood function was computed as $\mathbf{\Sigma_t} = \mathbf{s\Phi s}$. The system consists of 39 parameters. We use a maximum likelihood approach to find parameter estimates. The valuation of the multivariate normal density function was made using the package `mvtnorm` Genz et al. (2020) in the software language R (R Core Team, 2020).

4.2.4 Validation

The model validation consisted of testing the predictive ability of the system of models for an out-sample observation. We used the variance explained based on cross-validation ($VEcv$) (Equation (4.24)) recommended by Li (2017) as the best reliable accuracy measure for model validation. We also evaluated the residual mean $\bar{E}_i$ (Equation (4.25)), and the root mean square difference ($RMSD$) as proposed by Stage and Crookston (2007).

$$VEcv_i = \left(1 - \frac{\sum_{j=1}^{n}\sum_{t=1}^{T}(R_{ijt} - \hat{R}_{ijt})^2}{\sum_{j=1}^{n}\sum_{t=1}^{T}(R_{ijt} - \bar{R}_{ijt})^2}\right) 100 \tag{4.24}$$

$$\bar{E}_i = \frac{\sum_{j=1}^{n}\sum_{t=1}^{T}(R_{ijt} - \hat{R}_{ijt})}{nT} \tag{4.25}$$

$$RMSD_i = \sqrt{\frac{\sum_{j=1}^{n}\sum_{t=1}^{T}(R_{ijt} - \hat{R}_{ijt})^2}{nT}} \tag{4.26}$$

where R_{ijt} and $\hat{R}_{ijt}$ is the observed and predicted value for the ith response with the jth plot at the time t, n is the number of plots, and T is the number of observations within a single measured plot.

4.3 Results

The parameter estimates for models (4.4), (4.7), and (4.16) are shown in Table 4.4. The approximate confidence intervals for all the parameters constructed with the bootstrap standard error indicates that almost all the main parameter and factors are important in the model.

The approximate 95% confidence interval for the estimate parameter $\hat{a}_6$ includes zero, which indicate that the effect of Upper Coastal Plain and Lower Coastal Plain in equation (4.2) is the same. The parameter b_2 presented large variability, and the 95% confidence interval included zero, which indicates that a more simple model could be obtained. However, to maintain the compatible growth and yield model conditions, we decide to keep this parameter in the model.

Although the MRT study design aggregates the UC and PI physiographic regions as one stratum, we modeled them as separate geographic regions. We found that for DH and BA the three physiographic regions exhibit important differences in parameter values (see Table 4.4).

The residual diagnostic plots and a comparison of observed and predicted values for each of the three state variables indicate that the simultaneous fitting successfully achieves an

Table 4.4: Parameters estimates for equations (4.4), (4.7), and (4.16). Boot SE = Bootstrap estimate standard error, Lower Limit, Upper Limit = lower and upper bound for an approximate 95% confidence interval.

Parameter	Estimate	Boot SE	Lower Limit	Upper Limit	Parameter	Estimate	Boot SE	Lower Limit	Upper Limit
$\hat{a}_0$	1.7811	0.0902	1.6038	1.9584	$\hat{d}_1$	20.7664	2.9640	14.9416	26.5912
$\hat{a}_1$	-2.742e-5	2.954e-7	-2.799e-5	-2.683e-5	$\hat{d}_2$	25.4754	3.7524	18.1012	32.8496
$\hat{a}_2$	-0.8526	0.2716	-1.3863	-0.3189	$\hat{d}_3$	14.1619	1.4066	11.3977	16.9261
$\hat{a}_3$	0.2284	0.0754	0.0801	0.3766	$\hat{b}_2$	0.1602	0.1014	-0.0391	0.3594
$\hat{a}_4$	0.0250	0.0066	0.0119	0.0380	$\hat{b}_{30}$	-0.3137	0.0577	-0.4271	-0.2002
$\hat{a}_5$	0.0150	0.0047	0.0058	0.0243	$\hat{b}_{31}$	-0.0352	0.0081	-0.0510	-0.0194
$\hat{a}_6$	0.1618	0.1033	-0.0412	0.3648	$\hat{b}_{32}$	-0.0275	0.0099	-0.0469	-0.0081
$\hat{a}_7$	-0.0936	0.0130	-0.1192	-0.0680	$\hat{\gamma}_1$	-4.2440	0.1455	-4.5298	-3.9581
$\hat{k}$	0.0439	0.0043	0.0355	0.0524	$\hat{\gamma}_2$	-0.6748	0.0596	-0.7920	-0.5576
$\hat{n}_1$	11.5935	1.8150	8.0267	15.1604	$\hat{\gamma}_3$	0.5051	0.0504	0.4061	0.6041
$\hat{n}_2$	4.0936	1.3153	1.5088	6.6783	$\hat{\gamma}_4$	4.8698	0.4565	3.9728	5.7668
$\hat{n}_{30}$	3.6310	0.5343	2.5809	4.6811	$\hat{\gamma}_5$	0.3226	0.0284	0.2667	0.3784
$\hat{n}_{31}$	-0.0198	0.0031	-0.0258	-0.0138	$\hat{\gamma}_6$	8.0178	0.8514	6.3447	9.6910
$\hat{b}_{01}$	0.1493	0.0690	0.0138	0.2849	$\hat{\gamma}_7$	0.2732	0.0290	0.2162	0.3302
$\hat{b}_{02}$	0.0370	0.0123	0.0128	0.0612	$\hat{\gamma}_8$	0.0950	0.0101	0.0752	0.1148
$\hat{b}_{03}$	0.0398	0.0193	0.0020	0.0777	$\hat{\gamma}_9$	0.0233	0.0019	0.0196	0.0270
$\hat{b}_{04}$	11.4625	3.9969	3.6080	19.3171	$\hat{\phi}_{DH\_N}$	-0.1528	0.0148	-0.1820	-1.1236
$\hat{c}_1$	0.0556	0.0119	0.0322	0.0790	$\hat{\phi}_{DH\_BA}$	0.2674	0.0224	0.2234	0.3114
$\hat{c}_2$	2.379e-8	1.096e-14	2.379e-8	2.379e-8	$\hat{\phi}_{N\_BA}$	0.2172	0.0211	0.1756	0.2587
$\hat{c}_3$	0.1047	0.0252	0.0551	0.1542					

adequate representation of system dynamics (see Figure 4.2). The residuals for DH and BA show acceptable error distributions and unbiased predictions. The stand density model's distribution of errors indicates that the non-thinned control plots' values, with the largest stand densities, were predicted with less precision than the thinned plots. This situation is also reflected in the estimated parameters of γ_5 and γ_6. When the model (4.20) is used for non-thinned control plots, the standard deviation of error predictions is increased eight times for each increment unit in dominant height. For treated plots, the fitted stand density provides good predictions with lower errors. However, considering the wide range of plot densities included (from 247 trees ha^{-1} to 1735 trees ha^{-1}), the errors for treated plots with MRSP are relatively low.

The figure 4.3 presents the Weibull modifiers for Fert, CVC, and their combination. On average, the CVC treatment shows an immediate effect on the gross BA growth, and as the dominant height increases its effect decreases following a inverse-j shape. Fert only produces

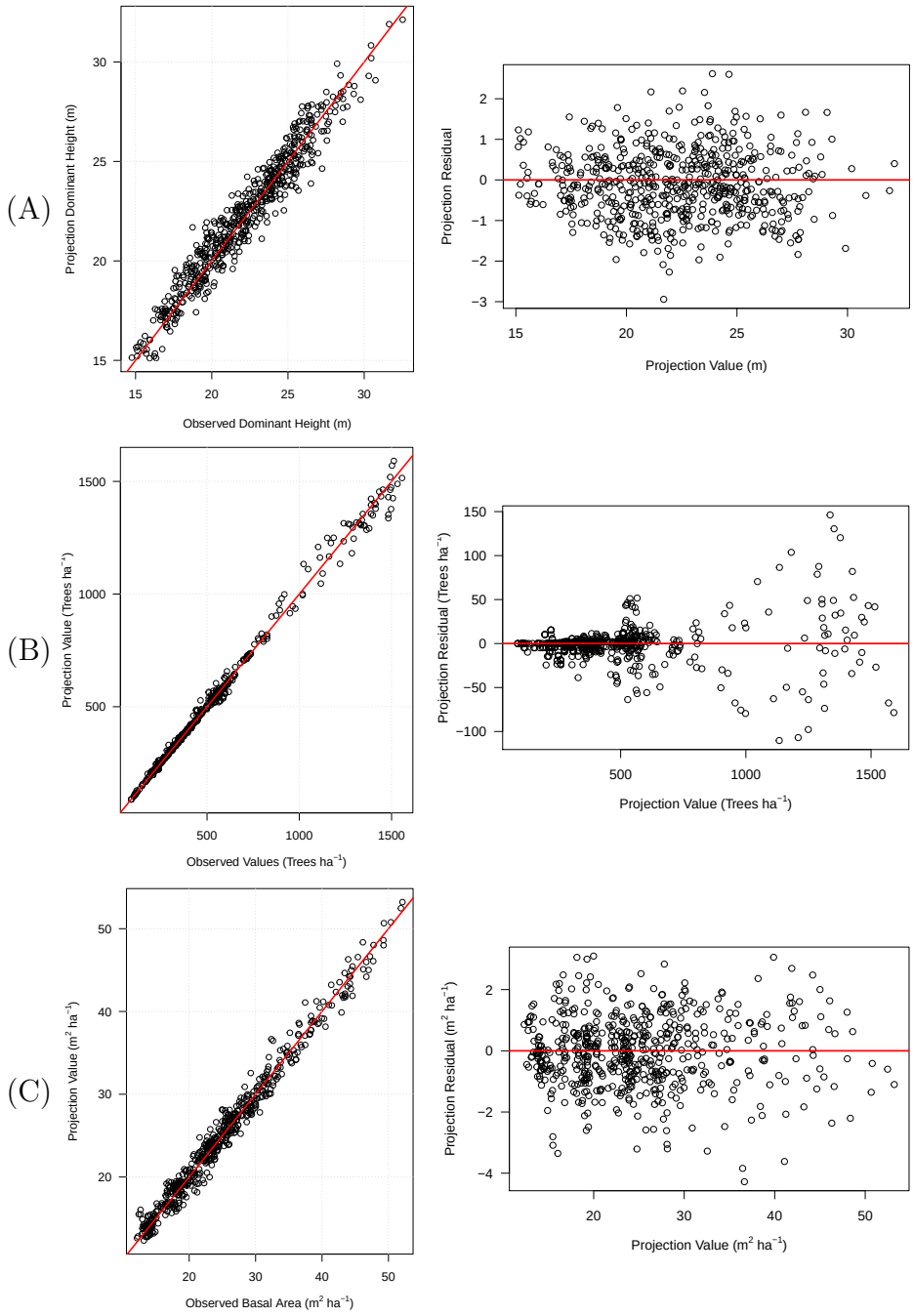

Figure 4.2: Observed and predictions for the three stand-level variables (left) and residuals compared with estimated values (right) for dominant height (A), stand density (B), and basal area (C). The diagonal red line has slope one and origin zero.

an increased growth response until the DH has gained 1.3 m after treatment. After that point, the effect starts to decrease following a similar pattern like CVC. The combinations of treatments Fert + CVC produces a large effect on the gross BA growth, but it is less than additive. The maximum value for the combined treatment was reached when the DH increment was exactly 1.8 m after treatment. After that point, the growth effect decreases with DH increases. In Figure 4.3 we extrapolate the DH increments until 25 m to see the behavior of the modifiers over a long run. The growth effect in the three treatments decreases, and they approach asymptotically to zero as was expected. The fertilization effect has a long-lasting effect on the sites. The combined treatment Fert+CVC produces an equal effect to CVC after a DH increment of 22.6 m.

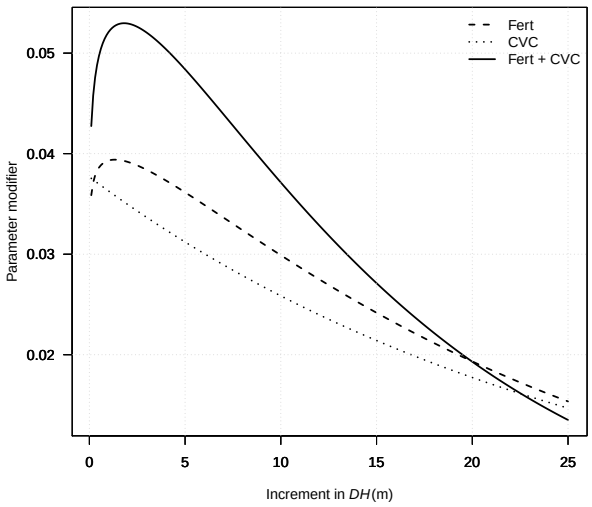

Figure 4.3: Magnitude and behavior of the modifier Weibull function for three mid-rotation silvicultural practices. Fertilization (Fert), Competitive vegetation control (CVC), and their combination (Fert + CVC). The modifier function is coupled with the basal area growth equation, Equation (4.13).

The estimated parameters for the projection standard deviation models for DH, N and BA in equations (4.19), (4.20), and (4.21) are also in Table 4.4. The predicted error standard deviations all exhibit an increasing trend as the time of projection increases (Figure 4.4). The slope of this uncertainty projection is notoriously different between LC and UC/PI. The

projection uncertainty for a non-thinned stand is always higher than for thinned stand (see Figure 4.4). As the estimated parameter $\hat{\gamma}_2 < 0$, then the projection standard deviation for DH increases at a logarithmic rate. The best model we found to express the projection uncertainty for N and BA resulted in nearly a linear trend (see Figure 4.4, and Equations 4.20, 4.21). This result is biologically sound given that Equations (4.5), (4.9), and (4.17) will project values in the future as a function only of the initial observed condition of the state variables at the treatment age.

4.3.1 Validation

Predictivity ability and performance for each response variable were evaluated using leave-one-plot-out cross-validation (243 plots). The results of the cross-validation are presented in Table (4.5). The cross-validation variance explained ($VEcv$) for each response variable using all the validation plot-measurements out-sample are 96.2% for dominant height, 99.7% for stand density, and 98.6% for basal area (Table 4.5). The residual mean ($\bar{E}$) is -0.04 m for dominant height, -0.73 trees ha^{-1} for stand density, and -0.075 m^2 ha^{-1} for basal area. In general, The small residual mean and high variance explained obtained with cross-validation indicate that this model system produces unbiased predictions. We are confident it will reflect the treatment effect across the southern physiographic regions when used for long-term projections. The non-thinning conditions produce the highest RMSD for stand density (37.9 trees ha^{-1}) and basal area (1.32 m^2 ha^{-1}) compared with others plots that have received thinning at least one time (treatments Thin, Fert, CVC, and Fert + CVC). This result is consistent with the differences in standard prediction error observed in Figure 4.4 between non-thinned control and first-thinned plots.

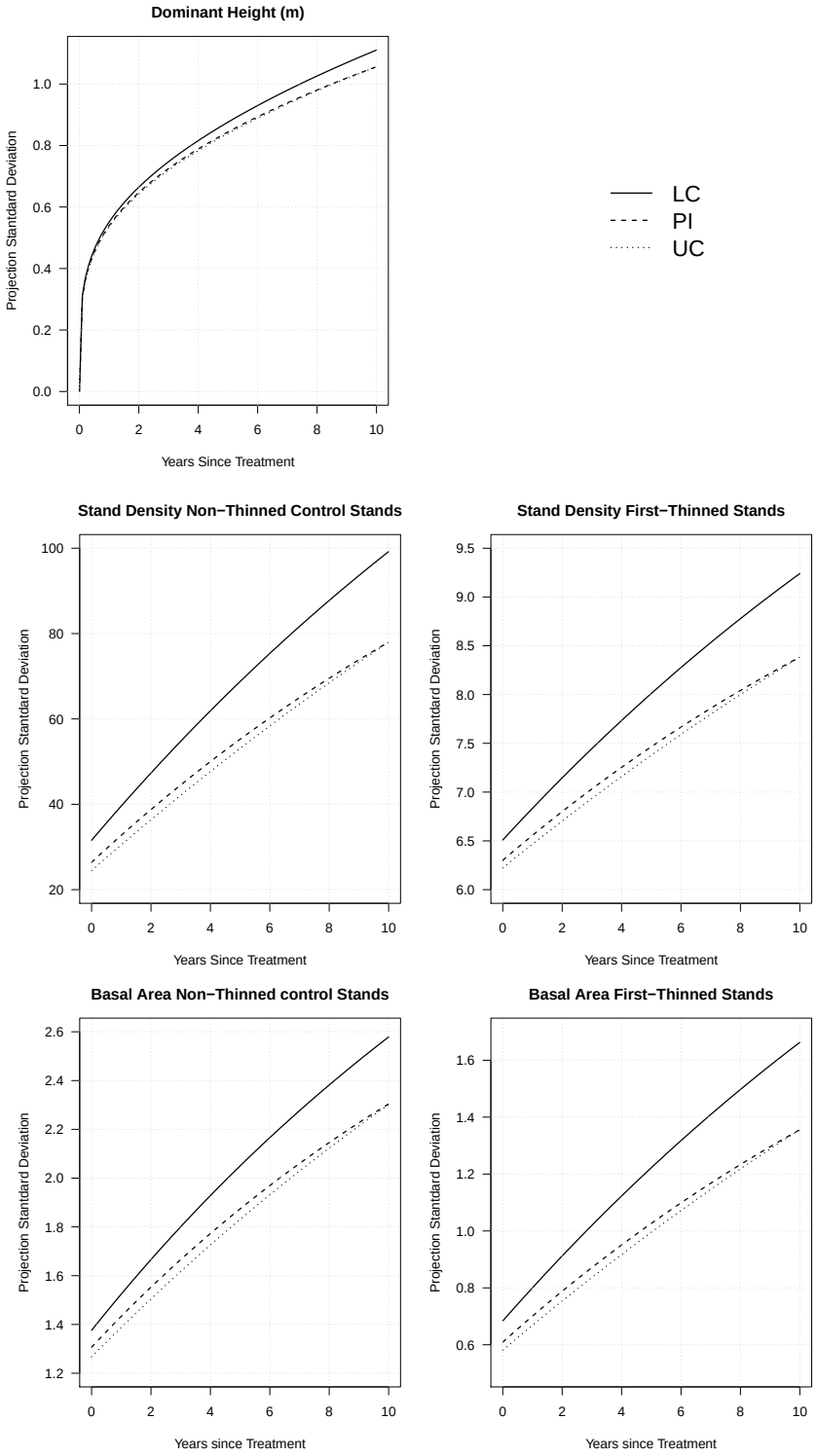

Figure 4.4: Projected error standard deviation for dominant height (m), stand density (trees ha^{-1}), and basal area (m^2 ha^{-1}) by physiographic region using Equations (4.19), (4.20), and (4.21). All projections were calculated assuming a loblolly pine plantation thinned at age 12 years, 1493 trees ha^{-1} before thinning, 489 trees ha^{-1} remaining after thinning, 11.5 m^2 ha^{-1} of basal area after thinning, 15.5 m of dominant heights at age 12, and a thinning intensity $R = 0.672$.

Table 4.5: Cross-validation variance explained (Equation (4.24)), prediction residual mean (Equation (4.25)), and prediction root mean square distance (RMSD) (Equation (4.26)) for each of the three response variables: dominant height (DH), stand density (N), and basal area (BA). **n** = number of plot-observation predicted out-sample.

Response	**Group/ Treatment**	**n**	**Variance Explained (%)**	**Residual Mean**	**RMSD**
DH (m)	LC	335	96.3	-0.0646	0.649
	PI	105	98.0	0.117	0.519
	UC	393	95.5	-0.0645	0.787
	Non-Thin	160	96.6	-0.0289	0.673
	Thin	162	96.1	-0.162	0.679
	Fert	169	95.9	-0.0421	0.751
	CVC	169	96.1	-0.0578	0.729
	Fert+CVC	173	96.4	0.0759	0.682
	All	833	96.2	-0.0416	0.704
N (trees ha^{-1})	LC	335	99.5	-2.07	23.4
	PI	105	99.7	4.73	19.9
	UC	393	99.8	-1.05	14.0
	Non-Thin	160	99.1	-1.25	37.9
	Thin	162	99.8	-0.566	7.67
	Fert	169	99.4	-0.866	12.1
	CVC	169	99.6	0.340	9.61
	Fert+CVC	173	99.5	-1.32	11.4
	All	833	99.7	-0.730	19.1
BA (m^2 ha^{-1})	LC	335	98.4	-0.0196	1.01
	PI	105	98.9	0.0483	0.907
	UC	393	98.6	-0.155	1.06
	Non-Thin	160	96.1	-0.140	1.32
	Thin	162	96.9	-0.0386	0.911
	Fert	169	97.0	-0.101	0.955
	CVC	169	97.2	-0.0404	0.924
	Fert+CVC	173	97.1	-0.0582	0.975
	All	833	98.6	-0.0751	1.03

4.3.2 Treatment Effects

Considering that the model system fits well, and the cross-validation results indicate an acceptable predictability ability, Equation (4.17) can be used for inference and to characterize the basal area response type due to the MRSP. Figure 4.5 shows the result of using the

Equation (4.5), (4.9), and (4.17) to simulate the response in basal area of a stand that was thinned by the first time at age 12, the basal area before thinning was 26.7 m^2 ha^{-1}, the number of trees before thinning was 1493 trees ha^{-1}, the thinning intensity $R = 0.672$, the remaining basal area was 11.5 m^2 ha^{-1}, the remaining number of trees was 489 trees ha^{-1}, and the dominant height at the moment of the thinning was 15.5 m (site index of 28.2 m at base age 25). The same simulated plot was used in each physiographic region to visualize the treatment effects across regions. All simulations were compared with the basal area of the thin only condition. In this particular example, after 20 years since treatments (age 32), all the treated plots (Fert, CVC, and Fert + CVC) respond positively to the treatments and result in significant differences compared with thin only (see Figure 4.5).

The combined treatment produces the largest mean response in the three regions. LC is the most responsive region in all the treatments. The BA response for treatment Fert+CVC was 6.4 m^2 ha^{-1}, which is 18.6 percent of the accumulated basal area with thin only. In this example, after 20 years, the treatment effects persisted in producing favorable growth rates in basal area. However, the large projected standard deviation obtained with the fitted model 20 years after treatment does not conclude statistical differences between treatments. The Weibull modifier influences the type of response observed in Figure 4.5. For all the installations available in the MRT study, we found that the combined treatment produces a less than additive response.

The basal area projection equation was used to compare the expected relative increment after 20 years since treatment in all the installations. The response pattern Fert + CVC > Fert > CVC was observed in 26 of 49 installations. Fert produced the larges response in 7 of 49 installations, CVC produced the largest response in 4 of 49 installations. The pattern Fert + CVC > CVC > Fert was observed in 8 of 49 installations. The Fert and CVC produced equal relative increments in 3 of 49 installations. In only one installation, all the treatments produced the same relative increment.

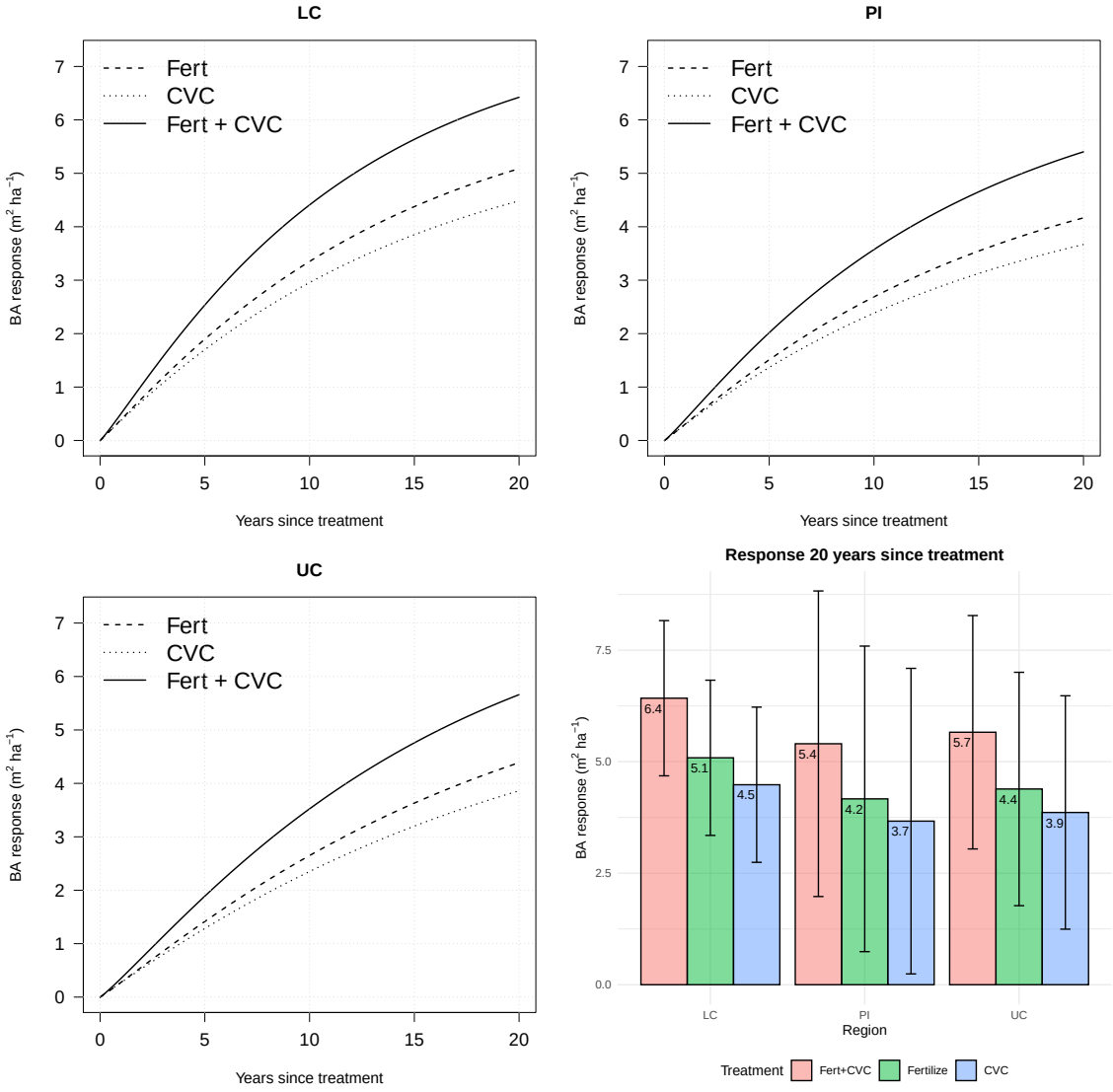

Figure 4.5: Basal area response of loblolly pine to mid-rotation treatments after thinning (Fert = fertilization, CVC = competitive vegetation control, and Fert + CVC = combination of fertilization and competitive vegetation control), for a simulated stand in three physiographic regions (LC = Lower Coastal Plain, UC = Upper Coastal Plain, and PI = Piedmont) in the southeastern U.S. The simulated stand received the first thinning at age 12, the basal area before thinning was 26.7 m^2 ha^{-1}, the number of trees before thinning was 1493 trees ha^{-1}, the thinning intensity $R = 0.672$, the remaining basal area was 11.5 m^2 ha^{-1}, the remaining number of trees was 489 trees ha^{-1}, and the dominant height at the moment of thinning was 15.5 m (site index 28.2 m at base age 25). The expected Basal Area responses 20 years since treatment by region (bottom left) are shown with its respective 95% confidence intervals.

4.4 Discussion

Different modeling assumptions and approaches have been proposed to mode MRSP effects. It ranges from empirical models at the tree level (e.g., Soderbergh and Ledermann (2003); Albaugh et al. (2017)), or the stand level (e.g., Franklin et al. (2009)) to ones based on process-based models (e.g., Bryars et al. (2013); Subedi and Fox (2016)). The model presented here mostly undertake an empirical approach at the stand level but were formulated using growth theory principles. The fitted model system was designed to project stand attributes like dominant height, stand density, and basal area. Initial stand information at the moment of the thinning and immediately after thinning is required for projections. They can be obtained directly from field measurements or can be simulated using existing non-thinning regional base-models and applying thinning intensity indices to derive the remaining trees and basal area per hectare.

From a modeling perspective, the growth and yield (G&Y) system fitted with the MRT study provides an understanding of how thinning practice interacts with other mid-rotation treatments commonly applied across the southeastern U.S. Traditionally, responses to mid-rotation treatments have been included in G&Y systems using modifier sub-models (Pienaar, 1997; Amateis, 2000; Gyawali and Burkhart, 2015). However, previous approaches do not produce compatible G&Y systems. Our approach includes a Weibull probability density function as a modifier on parameter's function accounting for the direct effect and interaction of mid-rotation treatments into the basal area growth model. The Weibull-modifier showed to be convenient, bound the magnitude of the response, include a temporal component, and allows us to derive closed forms for the basal area yield. The inclusion of bounding parameters or bounding modifiers has been shown to be effective in modeling plantations responses (Scolforo et al., 2020).

We interpreted the pdf shapes as how the stand responds to the treatment relative to the DH growth. Over time, the DH growth after treatment will indicate when the plantation reaches a limit of productivity gain as the treatment ameliorates resource constraints. Two crucial characteristics arise from this new modeling strategy. First, using the DH change as a driver for the treatment response effect means that the model can adjust the projected response accordingly with the site quality. Second, the DH growth is directly linked to the physiological growth response for BA.

The dynamic G&Y model presented only includes modifiers in the BA growth model. The statistical variable selection criteria used do not support the hypothesis that DH or N was affected by the presence of MRSP compared with the thinning only control. Other authors have reported similar results with mid-rotation treatments. For example, Albaugh et al. (2004) reported no changes in stem density in a mid-rotation controlled experiment with optimal and sustained treatment applications of fertilization and irrigation. They also observed that nine years after treatment, there was a continuous increase in height growth rate for treated plots, explained by the very poor initial nutrient availability of the study site. In the control plots, without thinning, the most competition-related mortality occurred in diameter classes below the plot mean, which is consistent with other reports of the thinning effect in literature (Jokela et al., 2004). Contrary to the control plots, on the treated plots, the thinning practice minimize mortality losses, and on those with an extra MRSP, the BA growth was enhanced compared with thinning only treatment.

The estimated parameters for the model in Equation (4.4) indicate that the anabolic component for the DH growth is affected by location (physiographic region) and intensity of thinning (R). Specifically, the results of estimated parameters for Equation (4.2) and (4.3) reveal that physiographic regions and thinning intensities are essential to define the asymptotic and growth rate behavior for DH. MRSP, other than thinning, does not show important contributions in the DH model estimation. This result is consistent with Gyawali

and Burkhart (2015), who found no effect of mid-rotation fertilization on the DH asymptotic parameter.

The treatment Fert + CVC was the most responsive across all installations. Analyzing the confidence intervals for BA projections from the simulated stand in Figure 4.5 indicates that after 2.8 years, the BA with Fert + CVC treatment differs from its control. Twenty years after treatment, all three treatment responses present significantly larger values. Additionally, the order Fert + CVC > Fert > CVC was observed in 53% of the installations when comparing their projected relative basal area gain. The projected response for the combination of treatments was always less than additive. The authors in Albaugh et al. (2012) reported contrasting results. They studied fertilization and vegetation control responses in a 2 × 2 factorial design replicated in a wide range of sites across the Southeastern (13 sites). They found that the results of the interaction of treatments indicate that the combined treatment effects were additive, only four sites had a less than additive response. They conclude that other limitations beyond nitrogen and phosphorus were being ameliorated by the vegetation control.

Across all the physiographic regions, we found that 65% of the installations the BA response with Fert was larger than with CVC. This result agrees with other reports for loblolly pine in the southeastern U.S. The authors in Allen et al. (2005) reported that at mid-rotation, the low level of available soil nutrients, principally N and P, were the more limiting factor to growth rather than water limitations. In 22% of the installations, we found that CVC treatments produce equal or larger relative responses than Fert only. This changing positive result can be explained by the low local availability of nutrients and low efficacy of Fert when there is no vegetation control. The importance of competition for soil nutrients between pines and understory plants has been well established. In nutrient-limited environments such as these, sustained control of understory competition can reduce

the nutrient deficiencies in a pine plantation similar to annual fertilizer applications (Jokela and Martin, 2000).

The projection system of equations proposed here assumes that the stand's initial conditions when applying the treatments are known. The MRT experiment design supports this modeling assumption. The installation plots were carefully selected (was not a random sample within stand), seeking nearly similar basal area values and dominant height. The treatments were randomly assigned to plot following the same thinning intensity level within the installation. The first measurement was simultaneous or close to the treatment application; therefore, the first measurement, at the difference of the posterior measurements, does not have expressed the treatment effect. The estimation method proposed here then assume the first measurements as fix and known and are the initial condition to solve the differential equations in 4.4, 4.7, and 4.16.

4.5 Conclusions

The modeling strategies followed in this research offer a biological and mathematically consistent framework to assess loblolly pine growth responses due to mid-rotation treatments, including thinning, fertilization, competitive vegetation control, and their combinations. The proposed model system is a dynamic compatible growth and yield system rooted in theoretical and biological principles. The model proved to have high predictive accuracy and can be used to project the current stand attributes following combinations of mid-rotation silvicultural practices with different thinning levels and intensities. The modifier incorporated in the growth model for the basal area includes a temporal effect of treatments. Because of the model structure, the response to treatments changes with location, age when applied, and dominant height growth.

We use the regional Mid-Rotation Treatment (MRT) study established by the Plantation Management Research Cooperative (PRMC) at the University of Georgia, which includes

a variety of site qualities and growth conditions in thinned and non-thinned conditions. Therefore, the proposed model well represents regional growth conditions in loblolly pine plantations.

References

Albaugh, T. J., Fox, T. R., Rubilar, R. A., Cook, R. L., Amateis, R. L., and Burkhart, H. E. (2017). Post-thinning density and fertilization affect Pinus taeda stand and individual tree growth. *Forest Ecology and Management*, 396:207–216.

Albaugh, T. J., Lee Allen, H., Dougherty, P. M., and Johnsen, K. H. (2004). Long term growth responses of loblolly pine to optimal nutrient and water resource availability. *Forest Ecology and Management*, 192:3–19.

Albaugh, T. J., Stape, J. L., Fox, T. R., Rubilar, R. A., and Allen, H. L. (2012). Midrotation Vegetation Control and Fertilization Response in Pinus taeda and Pinus elliottii across the Southeastern United States. *Southern Journal of Applied Forestry*, 36(1):44–53.

Allen, H., Fox, T., and Campbell, R. (2005). What is ahead for intensive pine plantation silviculture in the South? *Southern Journal of Applied Forestry*, 29(2):62–69.

Amateis, R. L. (2000). Modeling response to thinning in loblolly pine plantations. *Southern Journal of Applied Forestry*, 24(1):17–22.

Amateis, R. L., Liu, J., Ducey, M. J., and Lee Allen, H. (2000). Modeling response to midrotation nitrogen and phosphorus fertilization in Loblolly Pine plantations. *Southern Journal of Applied Forestry*, 24(4):207–212.

Ballard, T. (1984). A simple model for predicting stand volume growth response to fertilizer application. *Canadian Journal of Forest Research*, 14(1):661–665.

Bertalanffy, v. (1949). Problems of organic growth. *Nature*, 163:156–158.

Blanchard, P., Devaney, R., and Hall, G. (2012). *Differential equations*. Brooks/Cole, Cengage Learning, Boston, MA, fourth edition edition.

Bryars, C., Maier, C., Zhao, D., Kane, M., Borders, B., Will, R., and Teskey, R. (2013). Fixed physiological parameters in the 3-PG model produced accurate estimates of loblolly pine growth on sites in different geographic regions. *Forest Ecology and Management*, 289:501–514.

Cannell, M. and Dewar, R. (1994). Carbon allocation in trees: a review of concepts for modelling. *Advances in Ecological Research*, 25:59–104.

Clutter, J. (1963). Compatible growth and yield models for loblolly pine. *Forest Science*, 9:354–371.

Clutter, J. L. and Jones, E. (1980). Prediction of growth after thinning in old-field slash pine plantations. In *Forest Service Research Paper*, SE-217, page 19. Southern Forest Experiment Station.

Finto, A., Schimleck, L. R., Daniels, R. F., and Clark III, A. (2011). Effect of fertilization on growth and wood properties of thinned and unthinned midrotation loblolly pine (Pinus taeda L.) stands. *Southern Journal of Applied Forestry*, 35(3):142–147.

Fox, T. R., Eric, J. J., and Allen, L. H. (2004). The evolution of pine plantation silviculture in the southern United States. In *Southern Forest Science: Past, Present, and Future*, volume 075 of *Gen.Tech.Rep SRS-75.*, page 394. U.S. Department of Agriculture, Forest Service, Southern Research Station.

Fox, T. R., Lee Allen, H., Albaugh, T. J., Rubilar, R., and Carlson, C. A. (2007). Tree nutrition and forest fertilization of pine plantations in the Southern United States. *Southern Journal of Applied Forestry*, 31(1):5–11.

Franklin, O., Aoki, K., and Seidl, R. (2009). A generic model of thinning and stand density effects on forest growth, mortality and net increment. *Annals of Forest Science*, 66(8):815–815.

Gallagher, D. (2019). *Uncertainty estimaton in forest growth and yield systems*. Dissertation, University of Georgia, Athens GA.

Garcia, O. (1983). A stochastic differential equation model for the height growth of forest stands. *Biometrics*, 39(4):1059.

Garcia, O., Burkhart, H. E., and Amateis, R. L. (2011). A biologically-consistent stand growth model for loblolly pine in the Piedmont physiographic region, USA. *Forest Ecology and Management*, 262:2035–2041.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2020). *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.1-0.

Gyawali, N. and Burkhart, H. E. (2015). General response functions to silvicultural treatments in loblolly pine plantations. *Canadian Journal of Forest Research*, 45:252 – 265.

Huggett, R., Wear, D. N., Li, R., Coulston, J., and Liu, S. (2013). Forecasts of forest conditions. In *The Southern Forest Futures Project: technical report*, The Southern Forest Futures Project: technical report, pages 73–101. Wear, D.N., and J.G. Greis, Forest Service Southern Research Station: General Technical Report SRS-GTR-178.

Jokela, E. J., Dougherty, P. M., and Martin, T. A. (2004). Production dynamics of intensively managed loblolly pine stands in the southern United States: a synthesis of seven long-term experiments. *Forest Ecology and Management*, 192(1):117–130.

Jokela, E. J. and Martin, T. A. (2000). Effects of ontogeny and soil nutrient supply on production, allocation, and leaf area efficiency in loblolly and slash pine stands. *Canadian Journal of Forest Research*, 30(10):1511–1524.

Jokela, E. J., Martin, T. A., and Vogel, J. G. (2010). Twenty-five years of intensive forest management with Southern pines: important lessons learned. *Journal of Forestry*, 108:338–347.

Kreutz, C., Raue, A., Kaschek, D., and Timmer, J. (2013). Profile likelihood in systems biology. *FEBS Journal*, 280(11):2564–2571.

Li, J. (2017). Assessing the accuracy of predictive models for numerical data: Not r nor r2, why not? Then what? *PLOS ONE*, 12(8):e0183250.

Miller, J. H., Zutter, B. R., Zedaker, S. M., Edwards, M. B., and Newbold, R. A. (2003). Growth and yield relative to competition for loblolly pine plantations to midrotation—A southeastern United States regional study. *Southern Journal of Applied Forestry*, 27(4):237–252.

Pienaar, L. (1997). An approximation of basal area growth after thinning based on growth in unthinned plantations. *Forest Science*, 25(2):223–232.

Pienaar, L. and Turnbull, K. (1973). The chapmant-richards generalization of von bertalanffy's growth model for basal area growth and yield in even-age stands. *Forest Science*, 19(1):2–22.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Sampson, D. and Allen, H. (1999). Regional influences of soil available water-holding capacity and climate, and leaf area index on simulated loblolly pine productivity. *Forest Ecology and Management*, 124:1–12.

Scolforo, H. F., Montes, C., Cook, R. L., Lee Allen, H., Timothy, J., Rubilar, R., and Campoe, O. (2020). A New Approach for Modeling Volume Response from Mid-Rotation Fertilization of Pinus taeda L. Plantations. *Forests*, 11(6):646.

Sharma, M., Smith, M., Burkhart, H. E., and Amateis, R. L. (2006). Modeling the impact of thinning on height development of dominant and codominant loblolly pine trees. *Annals of Forest Science*, 63(4):349–354.

Soderbergh, I. and Ledermann, T. (2003). Algorithms for simulating thinning and harvesting in five European individual-tree growth simulators: a review. *Computers and Electronics in Agriculture*, 39(2):115–140.

Stage, A. and Crookston, N. (2007). Partitioning error components for accuracy-assessment of near-neighbor methods of imputation. *Forest Science*, 53(1):62–72.

Stanturf, J. A., Kellison, R. C., Broerman, F. S., and Jones, S. B. (2003). Productivity of southern pine plantations. Where are we and how did we get here? *Journal of Forestry*, pages 26–31.

Subedi, S. and Fox, T. R. (2016). Modeling repeated fertilizer response and one-time midrotation fertilizer response in loblolly pine plantations using FR in the 3-PG process model. *Forest Ecology and Management*, 380:90–99.

Wear, D. N. and Greis, J. G. (2012). The southern forest futures project: summary report. In *The southern forest futures project: summary report*, SRS-168, page 68. Forest Service Southern Research Station.

Weiskittel, A. R., editor (2011). *Forest growth and yield modeling*. Wiley, Hoboken, NJ.

4.6 Appendix

4.6.1 Derivation of Equation 4.5

The differential form of the DH model (equation 4.4) is a first-order differential equation

$$\frac{dDH}{dt} = \alpha DH^{\mathbf{m}} - kDH \tag{4.27}$$

We can solve the integral turning (4.27) as a Bernoulli equation. First let's define $V = DH^{1-\mathbf{m}}$, then $DH = V^{\frac{1}{1-\mathbf{m}}}$. Differentiation of DH with respect t gives,

$$\frac{dDH}{dt} = \frac{1}{1-m} V^{\frac{\mathbf{m}}{1-m}} \frac{dV}{dt} \tag{4.28}$$

Substituting (4.28) in (4.27), and substituting DH by this new transformation in terms of V gives,

$$\frac{1}{1-m} V^{\frac{\mathbf{m}}{1-m}} \frac{dV}{dt} = \alpha \left(V^{\frac{1}{1-\mathbf{m}}} \right)^{\mathbf{m}} - k \left(V^{\frac{1}{1-\mathbf{m}}} \right) \tag{4.29}$$

After some algebra to solve for $\frac{dV}{dt}$ we get,

$$\frac{dV}{dt} = (1-\mathbf{m})\alpha - k(1-\mathbf{m})V \tag{4.30}$$

Equation (4.30) is a linear differential equation. Let's define the integrating factor $\mu(t)$ as

$$\mu(t) = e^{\int k(1-\mathbf{m})dt} \tag{4.31}$$

Note that the first derivative of the integrating factor in (4.31) is,

$$\frac{d\mu(t)}{dt} = e^{\int k(1-\mathbf{m})dt} k(1-\mathbf{m}) = \mu(t)k(1-\mathbf{m}) \tag{4.32}$$

and its integral is,

$$\int \mu(t)dt = \frac{1}{k(1-\mathbf{m})} e^{k(1-\mathbf{m})t} + C \tag{4.33}$$

Multiplying everything in (4.30) by $\mu(t)$ and rearrange terms, we get,

$$\mu(t)\frac{dV}{dt} + \mu(t)k(1-\mathbf{m})V(t) = \mu(t)(1-\mathbf{m})\alpha \tag{4.34}$$

The left side of (4.34) becomes $(\mu(t)V)'$ by the product rule in calculus. Then (4.34) reduce to

$$(\mu(t)V)' = \mu(t)(1-\mathbf{m})\alpha \tag{4.35}$$

Integrating both sides of (4.35) and solving for DH gives,

$$\begin{aligned}
\mu(t)V &= (1-\mathbf{m})\alpha\frac{1}{k(1-\mathbf{m})}e^{k(1-\mathbf{m})t} + C \\
e^{k(1-\mathbf{m})t}DH^{1-\mathbf{m}} &= \frac{\alpha}{k}e^{k(1-\mathbf{m})t} + C \\
DH^{1-\mathbf{m}} &= \frac{\alpha}{k} + Ce^{-k(1-\mathbf{m})t} \\
DH &= \left[\frac{\alpha}{k} + Ce^{-k(1-\mathbf{m})t}\right]^{\frac{1}{1-\mathbf{m}}}
\end{aligned} \tag{4.36}$$

The invariant equation is obtained from (4.36) solving for the constant C:

$$C = \frac{DH^{1-\mathbf{m}} - \frac{\alpha}{k}}{e^{-k(1-\mathbf{m})t}} \tag{4.37}$$

Assuming known initial conditions, $DH(t_0) = DH_{th}$, and $t_0 = t_{th}$, associated at the moment of the thinning, then the constant term in (4.37) is known and fixed. Substituting (4.37) into (4.36) and using the initial values condition gives,

$$\begin{aligned}
DH &= \left[\frac{\alpha}{k} + \left[DH^{1-\mathbf{m}} - \frac{\alpha}{k}\right]e^{-k(1-\mathbf{m})t}e^{k(1-\mathbf{m})t_{th}}\right]^{\frac{1}{1-\mathbf{m}}} \\
DH &= \left[\frac{\alpha}{k} + \left[DH^{1-\mathbf{m}} - \frac{\alpha}{k}\right]e^{-k(1-\mathbf{m})[t-t_{th}]}\right]^{\frac{1}{1-\mathbf{m}}}
\end{aligned} \tag{4.38}$$

Equation (4.38) is a projection equation for DH. The initial condition values comes from the stand DH and age at the moment of thinning. Equation (4.38) correspond with the equation presented in (4.5).

4.6.2 Derivation of Equation (4.7)

The stand density differential equation is separable and has the following form

$$\frac{dN}{dDH} = -\left(\frac{n_1}{1e+16}\right) DH^{n_2} N^{\mathbf{n_3}} \tag{4.39}$$

Grouping terms, (4.39) can be written as

$$N^{-\mathbf{n_3}} dN = -n_1 DH^{n_2} dDH \tag{4.40}$$

Integrating both sides of (4.40) gives,

$$\begin{aligned} \int N^{-\mathbf{n_3}} dN &= \int -n_1 DH^{n_2} dDH \\ \frac{N^{1-\mathbf{n_3}}}{-\mathbf{n_3}+1} &= -n_1 \frac{DH^{n_2+1}}{n_2+1} + C \\ N^{1-\mathbf{n_3}} &= n_1 \frac{\mathbf{n_3}-1}{n_2+1} DH^{n_2+1} + C \end{aligned} \tag{4.41}$$

Taken (N_{th}, t_{th}), and DH_{th} as initial values, the resulting invariant is,

$$C = N_{th}^{1-\mathbf{n_3}} - n_1 \frac{\mathbf{n_3}-1}{n_2+1} DH_{th}^{n_2+1} \tag{4.42}$$

substituting (4.42) into (4.41) and solving for N, we get,

$$
\begin{aligned}
N^{1-\mathbf{n_3}} &= n_1 \frac{\mathbf{n_3}-1}{n_2+1} DH^{n_2+1} + N_{th}^{1-\mathbf{n_3}} - n_1 \frac{\mathbf{n_3}-1}{n_2+1} DH_{th}^{n_2+1} \\
N &= \left[N_{th}^{1-\mathbf{n_3}} + n_1 \frac{\mathbf{n_3}-1}{n_2+1} \left[DH^{n_2+1} - DH_{th}^{n_2+1} \right] \right]^{\frac{1}{1-\mathbf{n_3}}}
\end{aligned} \tag{4.43}
$$

(4.43) is the projection form for stand density given the known initial stand density (N_{th}) and dominant height (DH_{th}) immediately after thinning.

4.6.3 Derivation of Equation (4.17)

The stand basal area differential equation has the following form

$$
\frac{dBA}{dDH} = \mathbf{b_0} N^{\mathbf{b_3}(b_2-1)} BA^{b_2} - \mathbf{b_3} \frac{BA}{N} \frac{dN}{dDH} \tag{4.44}
$$

Let's $V = BA^{1-b_2}$, then $BA = V^{\frac{1}{1-b_2}}$. Taken derivatives of BA respect to DH:

$$
\frac{dBA}{dDH} = \frac{1}{1-b_2} V^{\frac{b_2}{1-b_2}} \frac{dV}{dDH} \tag{4.45}
$$

Substituting (4.45) in (4.44) and rearranging terms, we get,

$$
\frac{1}{1-b_2} V^{\frac{b_2}{1-b_2}} \frac{dV}{dDH} = \mathbf{b_0} N^{\mathbf{b_3}(b_2-1)} \left(V^{\frac{b_2}{1-b_2}} \right)^{b_2} - \mathbf{b_3} \frac{V^{\frac{1}{1-b_2}}}{N} \frac{dN}{dDH} \tag{4.46}
$$

$$
\frac{1}{1-b_2} \frac{dV}{dDH} = \mathbf{b_0} N^{\mathbf{b_3}(b_2-1)} - \mathbf{b_3} \frac{V}{N} \frac{dN}{dDH}
$$

$$
\frac{dV}{dDH} + \mathbf{b_3}(1-b_2) \frac{V}{N} \frac{dN}{dDH} = \mathbf{b_0} N^{\mathbf{b_3}(b_2-1)} (1-b_2) \tag{4.47}
$$

Equation (4.47) can be solved using the Bernoulli's integration rule. Let's define the integrating factor as

$$\mu(DH) = exp\left\{\int \mathbf{b_3}(1-b_2)\frac{1}{N}\frac{dN}{dDH}dDH\right\} = N^{\mathbf{b_3}(1-b_2)} \tag{4.48}$$

Multiplying everything in (4.47) by $\mu(DH)$ gives,

$$\mu(DH)\frac{dV}{dDH} + \mathbf{b_3}(1-b_2)\mu(DH)\frac{V}{N}\frac{dN}{dDH} = \mu(DH)\mathbf{b_0}N^{\mathbf{b_3}(b_2-1)}(1-b_2) \tag{4.49}$$

The left side of (4.49) becomes the product rule $(\mu(DH)V)'$. Integrating both sides of (4.49) gives,

$$\begin{aligned} \mu(DH)V &= \int \mu(DH)\mathbf{b_0}N^{\mathbf{b_3}(b_2-1)}(1-b_2)dDH \\ N^{\mathbf{b_3}(1-b_2)}BA^{1-b_2} + C_1 &= \int \mathbf{b_0}(1-b_2)dDH \end{aligned} \tag{4.50}$$

The integral on the right-hand side is resolved as follow:

$$\begin{aligned} \int \mathbf{b_0}(1-b_2)dDH \quad &= (1-b_2)\int b_{01} + b_{02}IUC + b_{03}IPIdDH + \\ &\frac{b_{04}(1-b_2)}{Age_{th}}(If + Ir + Ifr)\int \left(\frac{\mathbf{c}}{\mathbf{d}}\right)\left(\frac{DH - DH_{th}}{\mathbf{d}}\right)^{\mathbf{c}-1} exp\left\{\frac{DH - DH_{th}}{\mathbf{d}}\right\}^{\mathbf{c}} dDH \\ &= (1-b_2)(b_{01} + b_{02}IUC + b_{03}IPI)DH + \\ &\frac{b_{04}(1-b_2)}{Age_{th}}(If + Ir + Ifr)\left[1 - exp\left\{-\left(\frac{DH - DH_{th}}{\mathbf{d}}\right)\right\}^{\mathbf{c}}\right] + C_2 \end{aligned} \tag{4.51}$$

Let's take the observed stand attributes immediately after thinning as initial condition values, $DH = DH_{th}$, $N = N_{th}$, $BA = BA_{th}$. Doing this, the invariant for the BA differential equation is,

$$C = BA_{th}^{(1-b_2)} N^{\mathbf{b_3}(1-b_2)} - (1-b_2)(b_{01} + b_{02}IUC + b_{03}IPI)DH_{th} \tag{4.52}$$

Substituting (4.52) in (4.50) and solving for BA gives,

$$\begin{aligned} BA &= \Big[[b_{01} + b_{02}IUC + b_{03}IPI)DH + W] (1-b_2) N^{\mathbf{b_3}(b_2-1)} + \\ &\quad \Big[BA_{th}^{(1-b_2)} N^{\mathbf{b_3}(1-b_2)} - (1-b_2)(b_{01} + b_{02}IUC + b_{03}IPI)DH_{th} \Big] N^{\mathbf{b_3}(b_2-1)} \Big]^{\frac{1}{1-b_2}} \end{aligned} \tag{4.53}$$

where

$$W = \frac{b_{04}}{Age_{th}} (If + Ir + Ifr) \left[1 - exp \left\{ - \left(\frac{DH - DH_{th}}{\mathbf{d}} \right) \right\}^{\mathbf{c}} \right]$$

and rearranging terms we get,

$$\begin{aligned} BA &= \Big[BA_{th}^{(1-b_2)} \left(\frac{N_{th}}{N} \right)^{\mathbf{b_3}(1-b_2)} + [b_{01} + b_{02}IUC + b_{03}IPI] (1-b_2) N^{\mathbf{b_3}(b_2-1)} [DH - DH_{th}] + \\ &\quad W(1-b_2) N^{\mathbf{b_3}(b_2-1)} \Big]^{\frac{1}{1-b_2}} \end{aligned} \tag{4.54}$$

Equation (4.54) is the basal area projection equation as presented in (4.17).

4.6.4 Modifiers Equations

$$f_{\Delta BA} = (1 - IS_j) BA_{S_U, \text{age } j} - (1 - IS_i) BA_{S_U, \text{age } i} \tag{4.55}$$

where $f_{\Delta BA}$ is the multiplicative basal area modifier, $BA_{S_U,j}$ is the basal area per hectare of the non-thinned counterpart with the same age j, same site index, and N as the thinned plantation $BA_{S_T,j}$. $IS = \frac{BA_{S_U,j} - BA_{S_T,j}}{BA_{S_U,j}} = 1 - \frac{BA_{S_T,j}}{BA_{S_U,j}}$ is the index of suppression at the age i after thinning, and IS_j is the suppression index at age j, with $i < j$ Pienaar (1997).

$$BA_{S_T,\text{age } t_2} = BA_{S_U,\text{age } t_1} + (BA_a - BA_b) + ln(\alpha(Z+1))\gamma Z^{\beta} \tag{4.56}$$

where $BA_{S_T,\text{age } t_2}$ is the basal area of a thinned stand at age t_2, $BA_{S_U,\text{age } t_1}$ is the basal area of the a reference non-thinned stand at age t_1 obtained from the available yield and growth models. $BA_a - BA_b$ is the thinning intensity measured as the basal area after minus basal area before thinning. The term $ln(\alpha(Z+1))$ reflects an initial effect of suppression in the basal area's development immediately after thinning. The power term γZ^{β} captures the increasing growth of the thinning stand relative to the control after the initial growth suppression stage. Z is years since thinning, β is a parameter to be estimated, parameters α, and γ are functions of stand variables Amateis (2000).

$$N_{S_t,\text{ age } t_2} = N_{S_u,\text{ age } t_2} + (N_a - N_b)exp(-\alpha Z) \tag{4.57}$$

where $N_{S_t,\text{ age } t_2}$ is the number of trees of the thinned stand at age t_2 following thinning, $N_{S_u,\text{ age } t_2}$ is the number of trees of the corresponding non-thinned stand at age t_2. The difference $N_a - N_b$ is the change in the number of trees due to the thinning. α is a parameter that is a function of stand variables at the time of thinning Amateis (2000).

$$f_{res} = \begin{cases} 1 & Z > k \\ \left(\frac{BA_b}{BA_a}\right)^{\frac{r\left[-(Z)^2+k(Z)\right]}{t^2}} & 0 < Z \leq k \end{cases} \tag{4.58}$$

where f_{res} is the treatment-response function or modifier, t is the stand age, r is the rate parameter, k is the duration parameter for thinning effect in years, Z is an integer defining years since treatment, and all others terms are as previously defined Gyawali and Burkhart (2015).

$$N_{t_2} = \left[N_{t_1}^{\beta_1} + \beta_2 \left(\frac{BA_a}{BA_b}\right)\left(\frac{SI}{100}\right)^{\beta_3}(t_2^{\beta_4} - t_1^{\beta_4})\right]^{1/\beta_1} \tag{4.59}$$

where N_{t_2} is the trees per hectare after thinning at age t_2, SI is the stand site index, N_{t_1} is the current number of trees at age t_1, with $t_1 < t_2$, and $\beta_1, \ldots, \beta_4$ are parameters Gyawali and Burkhart (2015).

$$R = KTACZQ \tag{4.60}$$

where K is a constant that depend of the species, $T = 1 - exp(-at^b)$ is the time-effect, t is years elapsed since fertilization, $A = 1 - exp(-cr)$ is the fertilization effect, r is the amount of fertilizer nutrient applied, C is the composition factor which depend of the specie, $C = 1$ if it is a pure stand, $Z = gs + hs^2$ is the stocking factor, s is the stocking expressed as a percentage of normal stocking, $Q = i - jq$ is the site quality, q is the site index, a, b, c, g, h, i, and j are parameters.

$$f_{res} = \left[1 + [(\rho_0 + \rho_1 \mathrm{P})\mathrm{N}] \frac{1}{\beta} exp\left\{-\frac{Z}{\beta}\right\}\right] \tag{4.61}$$

where N is the nitrogen doses (kg ha^{-1}), P is an indicator variable (P=1 if fertilized with P, otherwise P = 0), Z is years since treatment, and ρ_0, ρ_1, and β are parameters.

$$f_{res} = \left[1 + a_0^{\mathrm{P}} S^{a_1} \left(1 - exp\left\{-a_2 \mathrm{N}\right\}\right) \left(1 - exp\left\{-(b_0 + b_1 RS)Z\right\}\right)\right] \tag{4.62}$$

where P is a indicator variable, P = 1 if phosphorous is applied, P = 0 if not, N is the level of nitrogen (kg ha^{-1}), S is the estimate site index in m at a base age of 25 years, RS is the relative spacing, $RS = \frac{\sqrt{\frac{10000}{N}}}{DH}$, a_0, a_1, a_2, b_0, and b_1 are parameters.

$$\begin{aligned} f_{res} &= \alpha Z^{\beta} exp\{\gamma Z\} \qquad (4.63) \\ \alpha &= (1 - exp\{b_1 \mathrm{N}\})(DH/10)^{b_2}(A/10)^{b_3}(S/10)^{b_4}(N/100)^{b_5} + (D_1 b_6) \\ \gamma &= b_7 + b_8 ln(1 + \mathrm{P}) \end{aligned}$$

where Z is years since treatment, N is the amount (kg ha^{-1}) of nitrogen applied, P is the amount (kg ha^{-1}) of phosphorous applied, DH is the dominant height at fertilization, A is the stand age in years at fertilization, S is the site index base age 25 yr, N is number of trees per hectare at fertilization, D_1 is an indicator variable, $D_1 = 1$ if the site is somewhat poorly, poorly or very poorly drained, $D_1 = 0$ otherwise, β, $b_1, \ldots, b_8$ are parameters.

$$\begin{aligned} f_{res} &= \alpha Z^{\beta} exp\{\gamma Z\} \\ \alpha &= (1 - exp\{b_1 \mathrm{N}\})(DH/10)^{b_2}(N/100)^{b_3}(BA/10)^{b_4} \\ \gamma &= b_5 + (D_1 b_6 + D_2 b_7 + b_8) ln(1 + \mathrm{P}) \end{aligned} \tag{4.64}$$

where BA is the stand basal area at fertilization, D_1 is an indicator variable, $D_1 = 1$ if the site is very poorly drained, $D_1 = 0$ otherwise, D_2 is another indicator variable, $D_2 = 1$ for somewhat excessively, well, and moderately well drained sites, $D_2 = 0$ otherwise, $b_1, \ldots, b_8$ are parameters.

Chapter 5

Stand Diameter Distribution Projection Using Copulas[1]

[1]Zapata-Cuartas, M., B.P. Bullock, and C.R. Montes. Stand diameter distribution projection using copulas. To be submitted to *Forest*.

Abstract

A new procedure for predicting future stand diameter distribution given the main stand-level attribute projections and the initial diameter distribution was developed based on copulas of survival diameter distributions at two points in time. A normal copula with marginal three-parameter Weibull distribution was considered. A method to recover the future marginal Weibull diameter distribution compatible with stand-level projected attributes was presented. The results indicated that the copula method produces similar and sometimes better results than the traditional parameter-recovery method in Reynolds et al.'s (1988) error-index values when comparing plots with different mid-rotation treatments. A notable improvement in accuracy was noticed with the copula method with prolonged projections periods (7 years in this study). The new method was used in combination with a growth and yield equation system to illustrate the effect of mid-rotation treatments on the availability of wood products at rotation age with a hypothetical loblolly pine plantation. Further exploration of diameter distribution modeling with copulas are needed. However, so far, the presented approach results in a method that proved to be consistent and robust. Foresters and biometricians will be the beneficiary of this new method to accurately predict wood resource availability by size-classes.

5.1 Introduction

Forest Biometricians use stand-level models to project important stand characteristics, like total volume per hectare, basal area, or trees per hectare over-rotation age to evaluate productivity or expected financial returns. Often, the projected whole-stand structure needs to be described by a diameter distribution to produce low-level details and determine wood product composition and value. For example, to know the number of trees and its volume by diameter size-classes suitable for pulpwood, chip-n-saw, or sawtimber. The current procedure

used in forest modeling consists of finding a continuous probability density function (pdf) representing the stand's trees, and thus, the trees' relative frequency by diameter size-classes. Additionally, the overall stand yield should match the aggregate yield in each diameter class. There are good reasons behind this stand structure modeling. First, the diameter distribution model produces size-class information and describes stand structure, returning yield, and wood product availability (Bailey, 1973). Second, variables such as growth, volume, market value, conversion-cost, treatment responses, and future forest prescriptions are dependent on the tree's diameter distribution. Thus, stand management decisions can be evaluated and supported throughout reliable pdf simulations.

The objectives of this chapter were to

- To present a literature review on the use of pdfs to represent diameter distributions,
- To summarize the most common methods used to project diameter distributions over time, and
- To propose a new alternative method of diameter distribution projection over the time based on copulas and use the Mid-Rotation Treatment study (MRT) established by the Plantation Management Research Cooperative (PMRC) at the University of Georgia, described in Chapter 4, to compare with the traditional parameter-recovery method.
- To evaluate the effect of mid-rotation treatments on the availability of wood products at the rotation age using the new diameter projection method.

5.1.1 Diameter Distributions and pdf

Empirical evidence has shown that pdfs can fit well stand diameter distribution, and their use has a relatively long tradition in even-age growth modeling (Álvarez González et al., 2002). Specifically, for planted even-aged stands, it is preferred that distributions have a unimodal shape. Continuous probability distributions are the most used to characterize

stand diameters distributions (SDD). Although there are various theoretical pdfs, two of the most cited are the Weibull pdf (Cao, 2006; Bullock and Burkhart, 2005; Álvarez González et al., 2002) and Johnson's pdf (Mateus and Tomé, 2011), also known as Johnson's S_B (system bounded). This paper will review the Weibull distribution principally.

Traditionally, the temporal analysis of SDD is performed using standard univariate pdf approaches (one dimension $d = 1$). Contrary to the multivariate approach ($d \geq 2$), the standard marginal univariate method may follow two fundamental paths. One consists of fitting a univariate distribution to the observed tree list data and then using empirical models (e.g., diameter growth, quantiles growth, tree mortality, or survival) to project the SDD and summarize future stand information per diameter class. The second one, the desired future distribution parameters, are recovery from stand-level attributes and the pdf's moment. Although only one paper (Knoebel and Burkhart, 1991) was found that applies multivariate methods in SDD projection, multivariate modeling has interesting properties, and its prediction would be more realistic.

5.1.2 About the Weibull Distribution

The Weibull pdf is a widely used statistical model in engineering (Luko, 1999). In the Southern, the Weibull pdf is the more frequently cited. This pdf was initially proposed by Waloddi Weibull (1887-1979) while studying the strength properties of materials. He proposed this pdf as a new theory based on probability laws to explain experimental results about breaking load on pieces of a particular class of ductile materials (see Weibull (1939)), hence the name of this distribution. Then, the Weibull pdf was used in statistics as one of the extreme value distributions. In the forestry field, the Weibull pdf was proposed as a SDD (two and three parameters) by Bailey (1973). Since that, the use of this pdf in forestry literature became popular because of its versatility, inclusive can cover unimodal, reversed-J shapes, and a full range of skewness. The parameters are easily related to shape

and location, and therefore, can be associated with the stand attributes (Bailey, 1973). The pdf for a three-parameter Weibull is

$$\begin{aligned} f(DBH|c,b,a) &= \frac{c}{b^c}(x-a)^{c-1}exp\left\{-\left(\frac{DBH-a}{b}\right)^c\right\} \text{ for } c>0, \text{ and } b>0. \quad (5.1) \\ &= 0 \quad \text{otherwise} \end{aligned}$$

Where c represents the shape parameter, b represents the scale parameter, and a is the localization parameter, which in practice is a diameter close to the $min\{DBH\}$. The parameter c could be related to the stand development over the time: for $1 < c < 3.6$, the density function is unimodal and positively skewed. When $c \sim 3.6$ the, Weibull approximates a normal distribution (juvenile stands). As c increases above 3.6, the distribution becomes progressively more negatively skewed (non-thinned adult stands). The Weibull pdf has a close integrated form. The cumulative distribution function CDF (F) of the Weibull is obtained by integrating pdf using *change of variable* methods. The respective cumulative function is:

$$\begin{aligned} F(DBH|c,b,a) &= 1-exp\left\{-\left(\frac{DBH-a}{b}\right)^c\right\} \quad a \leq DBH < \infty \quad (5.2) \\ &= 0 \quad \text{otherwise} \end{aligned}$$

In forest biometrics, $F(x)$ has direct applicability for the construct of stand tables. For example, assume the total number of trees per hectare (TPH) is known, and the Weibull CDF is given. The number of trees in a diameter class having midpoint x and width $2w$ is given by $TPH_x = TPH * F(x+w) - F(x-w)$. So, In terms of the Weibull CDF: $N_x = N * exp^{[-(x-w)/b]^c} - exp^{[-(x+w)/b]^c}$. If the interest is to determine a particular diameter

x such that $F(x)\%$ of all the trees are smaller than x, then we can solve for x from the CDF: $x = b * \left[-ln(1 - (F(x)\%/100))^{1/c}\right]$.

For the two-parameter Weibull distribution, when the diameter x is equal to the scale parameter, the CDF reaches approximately the 63 percentile. Equivalently, in the three-parameter Weibull distribution, it happens when $x = a + b$.

5.1.3 Weibull's Parameters Estimation

Different methods have been reported to estimate parameters of pdf and CDF. Weibull pdf parameters can be fitted by the method of moments (Frazier, 1981) or the percentile method (Johns and Lieberman, 1966). Those methods were used due to easy implementation, low computational cost, and the estimators have some properties close to the efficient Maximum Likelihood estimators (ML).

Percentile methods consist of choosing two diameter percentiles from a regular sample (non-truncated) and create a system of two equations using the CDF, then solve for parameters b or c. Cohen (1965) and Harter and Moore (1965) explained the Maximum Likelihood estimation method for the Weibull distribution under scenarios of complete and censored data[2]. This method was used by Bailey (1973) to demonstrate the versatility of the Weibull as SDD. Recent literature uses the ML methods as a standard method for parameter fitting. Assuming $x_i \overset{iid}{\sim}$ Weibull(c, b, a), with the probability given by its density function in (5.1), the likelihood function of parameters given the data is:

[2]These two papers were carried out independently of the other, but they were published in the same journal volume in 1965. Dr. Cohen was a professor at the University of Georgia

$$
\begin{aligned}
L(c,b,a|\boldsymbol{x}) &= \prod_{i=1}^{n} f(x_i|c,b,a) \\
&= \prod_{i=1}^{n} \left[\frac{c}{b^c}(x_i-a)^{c-1} exp\left\{ -\left(\frac{x_i-a}{b} \right)^c \right\} \right] \\
&= \left(\frac{c}{\beta^c} \right)^n \prod_{i=1}^{n} (x_i-a)^{c-1} exp\left\{ -\left(\frac{\sum(x_i-a)}{b} \right)^c \right\} \qquad (5.3)
\end{aligned}
$$

Parameters that make maximum (5.3), also minimize the negative of log transformation. Its easy to show that the log-transformation of (5.3) is $l(c,b,a|\boldsymbol{x})$, with

$$
l(c,b,a|\boldsymbol{x}) = nlog(c) - cnlog(b) + (c-1)\sum_{i=1}^{n} log(x_i-a) - \sum_{i=1}^{n} \left(\frac{x_i-a}{b} \right)^c \qquad (5.4)
$$

Estimate parameters that minimize $-l(c,b,a|\boldsymbol{x})$ can be obtained iteratively using optimization algorithms.

Alternatively, the Weibull parameters can be estimated using the distribution moments. Although this method is not efficient compared with ML, its simplicity is used in SDD analysis. The parameters are estimated by computing the sample moments and setting them equal to the theoretical moments from the moment generating function. The moment generating function ($M_K(x)$) for the Weibull is as follow:

$$
M_K(x) = b^k \Gamma\left(1 + \frac{k}{c} \right) \qquad (5.5)
$$

where k represents the k^{th} theoretical moment, $\Gamma(.)$ is the gamma function, $\Gamma(\alpha) = \int_0^{\inf} x^{\alpha-1} exp^{-x} dx$, where $\alpha > 0$. Only the first two moments are needed to compute the parameter estimates for the two-parameters Weibull. The following identities are particularly useful

$$\begin{aligned} \frac{\frac{1}{n}\sum_{i=1}^{n} x_i^2}{\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right]^2} &= \frac{\Gamma\left(1+\frac{2}{\hat{c}}\right)}{\Gamma^2\left(1+\frac{1}{\hat{c}}\right)} \\ \hat{b} &= \frac{\frac{1}{n}\sum_{i=1}^{n} x_i}{\Gamma\left(1+\frac{1}{\hat{c}}\right)} \end{aligned} \tag{5.6}$$

The estimated shape parameter $\hat{c}$ can be solved first using root-finding techniques, then the estimated scale parameter $\hat{b}$ can be found from the last expression of (5.6).

Variations of the Moment-estimations methods include the use of predicted percentiles. For example, Mctague and Bailey (1987) used the 10th, 63th, and 90th percentile in a two-stage recursive technique in determining the compatible Weibull parameters. They used the identity $b = D_{63} + a$ to find an initial unadjusted parameter b^*, and the c parameter was recovered from the definition for 90th percentile using the Weibull CDF:

$$\begin{aligned} D_{90} &= a + b\left[-ln(1-0.9)\right]^{\frac{1}{c}} \\ c &= \frac{0.834032445}{ln\left[\frac{D_{90}-a}{b^*}\right]} \end{aligned} \tag{5.7}$$

With the estimates for a and c, the adjusted parameter b was estimated from Equation (5.6).

5.1.4 Model Evaluation

Some of the most used goodness-of-fit tests to corroborate if the marginal distribution were satisfactory or not is the Kolmogorov-Smirnov test (Duan et al., 2013; Binoti et al., 2012; Mateus and Tomé, 2011; Schreuder and Hafley, 1977), Anderson-Darling, and Cramer-von Mises test (Bullock and Burkhart, 2005). Other researchers have misused the chi-square statistic for this purpose because of dependency resulting from data grouping in classes (see

Cao (2006)). However, for reasons described extensively in Reynolds et al. (1988), these tests can be wildly inaccurate (for samples fewer than 40 trees per plot). Instead, they proposed an index based upon absolute deviations in the units of the response.

5.1.5 Projections of Diameter Distributions

Assume that certain stand-level information (instantaneous information) at time t_0, retrieved from the forest inventories, is given. Consider that the basal area (BA) and mean quadratic diameter (Dq) are known. The problem is to estimate the SDD pdf's parameters for any time t_2, $t_2 > t_0$, in such a way that the resulting predicted SDD is compatible with direct stand yield model predictions at t_2 (Burkhart and Tomé, 2012).

We can classify the existing and published methods in two groups: Parameter Prediction methods (PP) and Parameter Recovery methods (PR). In parameter prediction, stand-level information are used as covariates in developing regression methods to predict the distribution's parameters or the diameter percentiles. Several authors (Burkhart and Tomé, 2012; Cao, 2004) have found that PP has not been entirely satisfactory to project distributions. Typically lack of compatibility arises between the predictions of stand attributes from PP and those obtained with direct stand yield model projections (Burkhart and Tomé, 2012).

PR finds the distribution's parameters in such a way that compatibility between the whole-stand characteristics and those obtained through the theoretical distribution at the age of prediction is assured. In this method, percentiles or moments of the empirical diameter distributions are predicted as a function of stand attributes, e.g., age, basal area, or stand density. Theoretical distribution parameters are then recovered by leveraging the known relationships between the predicted attributes and the distributional moments.

Parameter recovery with a two-parameter Weibull pdf on the 25th and 97th percentiles was used by Bullock and Burkhart (2005) to characterize juvenile loblolly pine plantations with different spacing in Virginia and North Carolina. Duan et al. (2013) used percentiles

33rd and 90th for Chinese fir plantations in southern China. Breidenbach et al. (2008) used a generalized linear model (GLM) to estimate theoretical distributions' parameters using covariates from airborne Lidar metrics.

Cao (2004) compared six methods, including 1) Direct prediction: predict parameters directly from stand attributes using regression methods. 2) Moment estimation: use the first two moments of the diameter distribution to recover shape and scale. 3) Percentiles: predict parameters from the predicted quadratic mean diameter, minimum diameter, 25th, 50th, and 95th percentile. 4) A hybrid method: the parameters are recovered from moments using the predicted Dq and the 93rd percentile. 5) MLE: scale and shape parameters are expressed as functions of stand attributes, and the function's parameters are found iteratively to maximize the Weibull pdf log-likelihood. 6) CDF Regression: similar to 5), but in this case, the solution minimizes the sum of squares of errors for CDF. The author conclude that the method 6), CDF Regression, was superior and consistently outperformed the others.

5.1.6 Modeling Bivariate Distributions With Copulas

A copula is a function that connects or couples d marginal cumulative distribution functions $F_1, \ldots, F_d$ to the multivariate cumulative distribution function (*mcdf*) $\boldsymbol{H}$. In other words, the copula function determines the dependence between the components of d random vectors, each with *cdf* $F_1, \ldots, F_d$. Let us consider the case when $d = 2$. Assume that n sample observations of two continuous random variables $(\boldsymbol{X}_1, \boldsymbol{X}_2)$: $x_{1j}, \ldots, x_{nj}$, $j \in \{1, 2\}$ were observed. The joint cumulative distribution function (*mcdf*) $\boldsymbol{H}$ is defined as $\boldsymbol{H}(\boldsymbol{x}) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2)$, $x \in \mathbb{R}$ (where $\mathbb{P}$ denotes the underlying probability measure). The original definition of the copula is based on uniform univariate marginal $(0, 1)$. Most of the copula theory use uniform marginal for simplicity. Any continuous parametric pdf can be transformed to an equivalent uniform distribution through transformations. However,

in practical applications it is desirable to work directly with the parametric version of the marginals.

Hofert et al. (2018) presents two lemmas about two useful transformations. The Probability Integral Transformation (PIT) and the quantile transformation. Let us define F_j as a PIT such that $F_j(X)$, $j \in \{1, 2\}$ have standard uniform distribution $U(0, 1)$.

The copulas of (X_1, X_2) are the *cdf* of $(F_1(X_1), F_2(X_2))$. Equivalently, the *mcdf* $\boldsymbol{H}$ can be expressed as

$$\boldsymbol{H}(\boldsymbol{x}) = C\left(F_1(x_1), F_2(x_2)\right), \boldsymbol{x} \in \mathbb{R}^2, \tag{5.8}$$

The estimation of $\boldsymbol{H}$ using copulas allows one to model the marginals F_1, F_2 separately from the dependence structure represented by the copula C. The Sklar's theorem (Hofert et al., 2018; Owzar and Sen, 2013) supports the existence of a d-dimensional copula C.

Consider the case that $\boldsymbol{H}$, C, $\mathbf{F}_1$, and $\mathbf{F}_2$ are continuous. Then the density h of $\boldsymbol{H}$ satisfies:

$$h(\mathbf{x}) = \acute{c}(\mathbf{F}_1(x_1), \mathbf{F}_2(x_2)) \prod_{j=1}^{d=2} f_j(x_j), \quad \mathbf{x} \in \prod_{j=1}^{d=2} \text{ran} X_j, \tag{5.9}$$

where, ran $X_j = \{x \in \mathbb{R} : f_j(x) > 0\}$ is the range of the random variable X_j, f_j is the density of $\mathbf{F}_j$, and $\acute{c}$ is the density of C. The density of C can be recovered from h as

$$\acute{c}(\mathbf{u}) = \frac{h(\mathbf{F}_1^{\leftarrow}(u_1), \mathbf{F}_2^{\leftarrow}(u_2))}{f_1(\mathbf{F}_1^{\leftarrow}(u_1)) \times f_2(\mathbf{F}_2^{\leftarrow}(u_2))} \tag{5.10}$$

where $\mathbf{F}_j^{\leftarrow}()$ is the quantile transformation. u_1 and u_2 are random uniform (0,1) observations.

5.1.7 Parametric Copulas

Extensive options of copulas have been proposed in the literature. Some of them are defined as parametric copula families. For example, Clayton copula, Gumbel-Hougaard copula, Normal copula, Studnet t copula, and Galambos copula. Let us introduce some of them.

The Clayton family copula for $d = 2$ and parameter $\theta \in (0, \inf)$, are defined by

$$C_\theta^{\mathrm{C}}(\boldsymbol{u}) = \left(1 - d + \sum_{j=1}^{d} u_j^{-\theta}\right)^{-1/\theta}, \ \boldsymbol{u} \in [0,1]^d. \tag{5.11}$$

The larger the value of θ, the stronger the (positive) dependence between the components of $\boldsymbol{U} \sim C_\theta^{\mathrm{C}}$

The Gumbel-Hougaard family of copula for $d = 2$ and parameter $\theta \in [1, \inf)$, are defined by

$$C_\theta^{\mathrm{GH}}(\boldsymbol{u}) = exp\left\{-\left(\sum_{j=1}^{d}(-ln(u_j))^\theta\right)^{1/\theta}\right\}, \ \boldsymbol{u} \in [0,1]^d. \tag{5.12}$$

The Normal Copula for $d = 2$ is defined by

$$C_{\boldsymbol{\theta}}^{N}(\boldsymbol{u}) = \Phi_{\boldsymbol{\theta}}(\Phi^{-1}(u_1), \Phi^{-1}(u_2)), \tag{5.13}$$

where Φ^{-1} is the inverse cumulative distribution function of a standard normal distribution, $\boldsymbol{\theta}$ is the linear correlation matrix, and $\Phi_{\boldsymbol{\theta}}$ is the joint cumulative distribution function of a multivariate normal distribution.

5.1.8 Conditional Copulas

Let $\boldsymbol{Z}$ be q-dimensional auxiliary information matrix. In this paper, we are interested in the conditional *pdf* of $\boldsymbol{X}$ given $\boldsymbol{Z}$. The conditional cumulative distribution of X_1 and X_2 is

$$H_{\mathbf{z}}(\mathbf{x}) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2 | \mathbf{Z} = z), \quad \mathbf{x} \in \mathbb{R}^2. \tag{5.14}$$

Because $H_{\mathbf{z}}(\mathbf{x})$ is continuous, there exists a copula $C_{\mathbf{Z}}(F_{\mathbf{Z},1}(x_1), F_{\mathbf{Z},2}(x_2))$ where $F_{\mathbf{Z},j}(x_j) = \mathbb{P}(X_j \leq x | \mathbf{Z} = z)$, $j \in \{1, 2\}$ and $C_{\mathbf{Z}}$ is the conditional copula. Hofert et al. (2018) discussed the concept of conditional copulas in the context of regression. The general idea is to expand the process of copula estimation into a regression procedure. In this context, marginal calibration functions and a copula calibration function defined up to a finite-dimensional parameter vector $\boldsymbol{\beta}_j$ of regression coefficients are introduced in the estimation process. Thus the copula's parameter can be expressed in terms of the other auxiliary variables $z \in \mathbf{Z}$ in a regression setting. The method proposed in this study uses calibration functions to fit the conditional copula and estimate one of the marginals with the projected stand attributes.

5.2 Methods

5.2.1 Linking SDD and Copulas

Let us assume that k repeated measurements of SDD_t $(t = 0, \ldots, k)$ post-thinning are available for a permanent plot. In each measurement only the survival trees are tailed. SDD_0 is the observed SDD of surviving trees immediately after thinning and it is indexed with zero indicating that it is a base reference to be used to find a projection at age t $(t > t_0)$. We can also assume that the complete n pair of diameter observations (x_{i0}, x_{it}) $(i = 1, \ldots, n)$ are a random samples that belong from the an unknown survival bi-variate *mcdf* $\mathbf{H_t}$ between SDD_0 and SDD_t. Therfore, SDD_0 and SDD_t are the *pdf* marginal of $\mathbf{H_t}$. In this study we assume that the three-parameters Weibull distribution is appropiate as marginal *pdf*, however the conceptual frame of copulas allows any combination of continuous *pdf* marginals.

Note that the relationship between SDD_0 and SDD_t is based on the surviving trees at the two points of time t_0 and t_t exclusively. Therefore, Sklar's Theorem can well be adapted in terms of diameter survival dependence after treatment.

Theorem 5.1. *(**Sklar's Theorem for Diameter Survival Dependence**)*

1. *For any bi-variate* $\mathbf{H}$ *with margins* $\mathbf{F}_0$, $\mathbf{F}_t$, *from the surviving trees at point time t=0 (imediately after a mid-rotation treatment) and from the surviving trees at point* $t > 0$ *after treatment, respectively, there exists a diameter survival copula* C^s *such that*

$$\mathbf{H} = C^s(\mathbf{F}_0(x_0), \mathbf{F}(x_t)), \quad \mathbf{x} \in \mathbb{R}^2. \tag{5.15}$$

The diameter survival copula C^s *is uniquely defined on* $\prod_{j=1}^{2} ran\mathbf{F}_j$ *and there given by*

$$C^s(\mathbf{u}) = \mathbf{H}(\mathbf{F}_0^{\leftarrow}(u_0), \mathbf{F}_t^{\leftarrow}(u_t)), \quad \mathbf{u} \in \prod_{j=1}^{2} ran\mathbf{F}_j. \tag{5.16}$$

2. *Conversely, given any 2-dimensional diameter survival copula* C^s *and univariate cumulative distributions of surviving trees* $\mathbf{F}_0$, *and* $\mathbf{F}_t$, $\mathbf{H}$ *defined by (5.15) is a bivariate mcdf with margins* $\mathbf{F}_0$, $\mathbf{F}_t$.

Given a diameter survival copula C^s $(d = 2)$ and one marginal Weibull *pdf* derived from known initial condition immediately after treatment, one can consider obtaining the marginal Weibull *pdf* at point t from the copula, which is the SDD of interest. To this end, let's C^s be a 2-dimensional diameter survival copula, let $\mathbf{U} \sim C^s$, and for $u_0 \in (0, 1)$ and $u_0 \sim$ uniform with density $c^{t_0}(u_0) \sim$ Weibull,

$$C^s_{t|t_0}(u_t|u_0) = \mathbb{P}(U_t \leq u_t | U_0 = u_0) = \frac{\int_0^{u_t} c^{(t_0,t)}(u_0, u_t)du}{c^{t_0}(u_0)} \tag{5.17}$$

is the *cdf* on $[0, 1]$ for the SDD at point time t. The right-hand side of (5.17) provides the conditional diameter survival probability (or the probability distribution of diameters

that will survive at time t). This definition of conditional *df* was used to evaluate the goodness-of-fit of different copulas families to the data and to recover the SDD at any time $t > t_0$.

5.2.2 Selection of Copula for Diameter Survival Dependence

The exploration and estimation of bivariate copulas was made with pseudo-observation of diameters. This distribution-free transformation allows for scaling each diameter list to the range $(0, 1)$. Given n random realizations $x_i = (x_{it_0}, x_{it})$ $i \in 1, ..., n$, for diameters at time points t_0 and t in a particular plot, the respective pseudo-observation transformation is defined via $u_{ij} = r_{ij}/(n + 1)$ for $i \in 1, ..., n$ and $j \in \{t_0, t\}$, where r_{ij} denotes the rank of x_{ij} among all $x_{.j}$.

Preliminary exploratory analysis of SDD indicated that the bivariate relationship of survival trees between point times t_0 and t is radially symmetric and exchangeable (similar distribution of the observations below and above the main diagonal), and highly tail dependent (see Figure 5.1 for one particular example).

The exchangeability property was tested for all plots and measurements of the MRT using the Genest et al. (2012)'s test and implemented in `copula` package (Marius Hofert and Martin Mächler, 2011). The statistic is (Hofert et al., 2018; Genest et al., 2012),

$$S_n^{\mathrm{ex}_C} = \int_{[0,1]^2} n \left(C_n(u_0, u_t) - C_n(u_t, u_0)\right)^2 dC_n(\mathbf{u}), \tag{5.18}$$

where C_n is a non-parametric version of the copula or also called empirical copula computed as

$$C_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^{n} 1(\mathbf{U}_{i,n} \leq \mathbf{u}) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{2} 1(U_{ij,n} \leq u_j), \quad \mathbf{u} \in [0, 1]^2, \tag{5.19}$$

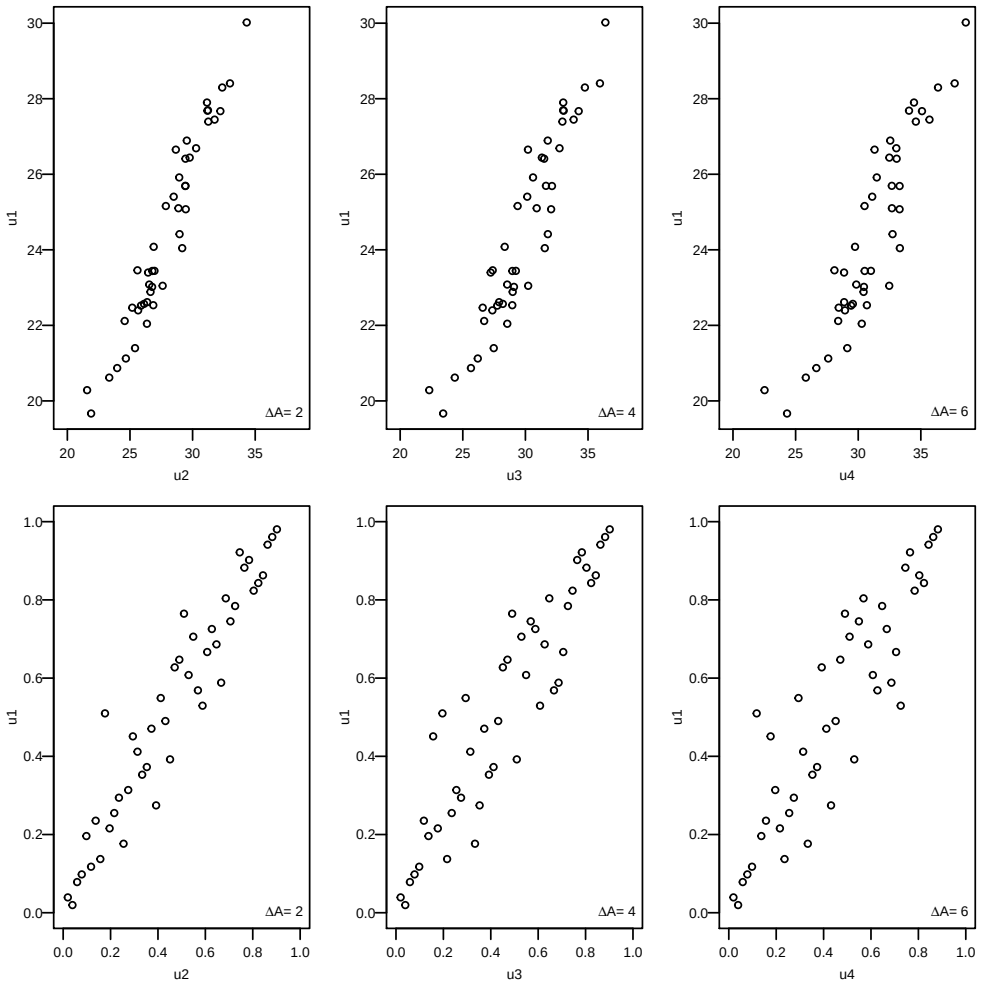

Figure 5.1: Scatter plots of diameters in centimeters (top row) and pseudo-observations (bottom row) for three post thinning measurements (u_2, u_3, u_4) of one permanent plot of the MRT data set. The time, in years since thinning, is indicated by $\Delta A = 2, 4, 6$. u_1 are the measurements immediately after thinning, u_2 are the measurements two years after treatment, u_3 are the measurements four years after thinning, and u_4 are the measurements six years after thinning.

where $\mathbf{U}_{i,n} = (U_{i0,n}, U_{it,n})$, $i \in 1, \ldots, n$, are the pseudo-observations and the logical inequalities $\mathbf{U}_{i,n} \leq \mathbf{u}$ are to be understood componentwise. Inference with the statistic in (5.18) are based on Monte-Carlo simulation.

The radial symmetry property was evaluated using the test proposed by Genest and Nešlehová (2014) and implemented in the `copula` package (Marius Hofert and Martin Mächler, 2011). The statistic is

$$S_n^{sym} = \int_{[0,1]^2} n \left(C_n(\mathbf{u}) - \bar{C}_n(\mathbf{u}) \right)^2, \tag{5.20}$$

where C_n is the empirical copula as defined in (5.19), and $\bar{C}_n$ is the empirical copula of $-\mathbf{X}_0, -\mathbf{X}_t$. Similar to the exchangeability test, inferences on (5.20) are based on a Monte-Carlo simulation with resampling on the observed data.

The Extreme-Value dependence property was evaluated using the test proposed by Ben Ghorbal et al. (2009). The statistics is (Hofert et al., 2018; Ben Ghorbal et al., 2009),

$$S_n^{ev} = -1 + \frac{8}{n(n-1)} \sum_{i \neq j} I_{ij} - \frac{9}{n(n-1)(n-2)} \sum_{i \neq j \neq k} I_{ij} I_{kj}, \tag{5.21}$$

where $I_{ij} = 1(X_{i0} \leq X_{j0}, X_{it} \leq X_{jt})$. The function `evTesK()` in the R package `copula` (Marius Hofert and Martin Mächler, 2011) provides an implementation of the test in (5.21).

The tests mentioned above are typically used to understand the type of dependence that is undertaken in the unknown copula C^s. We considered the exploration of copulas from the Archimedean class family like Clayton and Gumbel–Hougaard, from the Elliptical class family like Normal and (Student) t, and Extreme Value class family like Galambos copula. To assess whether the unknown copula C^s belongs to one of the types of chosen families, we performed a goodness-of-fit test. The null hypothesis is $\mathcal{H}_0 : C^s \in \mathcal{C}$, and the alternative is $\mathcal{H}_1 : C^s \notin \mathcal{C}$. Where $\mathcal{C}$ is one of the Archimedean, Elliptical, or Extreme-Value family. According to Hofert et al. (2018), one natural test consists of comparing the empirical

copula (Equation 5.19) with the fitted copula $\mathbf{C}_{\theta_n}$ under the null hypothesis. Genest et al. (2009) recommends a procedure for goodness-of-fit assessment based on the Cramér-von Mises statistics. For the bivariate case, the statistic is:

$$S_n^{gCMM} = \int_{[0,1]^2} n \left(C_n(\mathbf{u}) - C_{\theta_n}(\mathbf{u})\right)^2 dC_n(\mathbf{u}) = \sum_{i=1}^{n} \left(C_n(\mathbf{U}_{i,n}) - C_{\theta_n}(\mathbf{U}_{i,n})\right)^2 . \tag{5.22}$$

We used the implementation of the S_n^{gCMM} test available in the R package `copula` (Marius Hofert and Martin Mächler, 2011). Inferences are obtained by parametric bootstrap (1000 repetitions of pseudo-random samples for each plot-time comparison) as described by Hofert et al. (2018).

Finally, based on the results of goodness-of-fit, we select the copulas that showed the best performance fitting the survival diameter dependence and ranked them using the Copula Information Criterion (CIC) in the context of leave-one-out cross-validation as suggested by Hofert et al. (2018) and Grønneberg and Hjort (2014). The implementation of these statistics in the `copula` package returns the log-density of copula evaluate in each out-sample $\mathbf{X_i}$ with parameters estimated with the corresponding remaining data. By construction, the CIC statistic penalizes copulas families with too many parameters that tend to overfit. Therefore, the selection criteria consist of selecting the family that produces the largest CIC.

5.2.3 Projection of Stand Diameter Distribution

Next is the description of the method used to estimate a system of equations to project a know SDD from the point time t_0 to a future time t using copulas:

- Estimate the Weibull three-parameters pdf for the tree-list diameters at t_0 using the Maximum Likelihood method with the observed minimum diameter (D_{min,t_0}) fix. Using the estimated Weibull pdf parameters, find the pdf quantiles 10, 25, and 95%.

- Compute the mean square diameter for t_0 and t (Dq_{t_0} and Dq_t, respectively) from the known stand-level information of basal area and number of trees per hectare. Obtain the dominant height estimations for t_0 and t (DH_{t_0} and DH_t, respectively). The stand-level information at time t can be obtained from projections with the Growth and Yield equation system presented in Chapter 4.

- Obtain thinning information, if any, at time t_0 and calculate the Thinning Intensity (R) as the number of trees removed over the amount of the initial number of trees per hectare.

- The database used to fit the model include all the possible combinations of pairs of measurements (t_i, t_j), with $i < j$ within the plot.

- Predict the parameter of the parametric copula C_s with the following equation

$$\theta = exp\left\{-(\beta_1 + \beta_2 \frac{t_0}{100} + \beta_3 R)(t - t_0)\right\} \tag{5.23}$$

  The argument here is that the dependence defined by C^s is also a function of the time passed since t_0, and the stand condition after thinning (if any) expressed by the thinning intensity.

- Predict the quantile k ($q_{k,t}$) for $k = 25\%$ or 95% at point time t with the following equations

$$q_{25,t} = exp\left\{\left(\frac{t_0}{t}\right) ln(q_{25,t_0}) + \left(1 - \frac{t_0}{t}\right)(\beta_4 + \beta_5 ln(Dq_2) + \beta_6\theta + \beta_7 \frac{DH_{t_0}}{DH_t})\right\} \tag{5.24}$$

$$q_{95,t} = exp\left\{\left(\frac{t_0}{t}\right) ln(q_{95,t_0}) + \left(1 - \frac{t_0}{t}\right)(\beta_8 + \beta_9 ln(Dq_2) + \beta_{10}\theta + \beta_{11}\frac{DH_{t_0}}{DH_t})\right\} \quad (5.25)$$

- Predict the minimum diameter $D_{min,t}$ at time t with the following equation

$$D_{min,t} = exp\left\{\left(\frac{t_0}{t}\right) ln(D_{min,t_0}) + \left(1 - \frac{t_0}{t}\right)(\beta_{12} + \beta_{13} ln(q_{10,t_0}) + \beta_{14}\theta)\right\} \quad (5.26)$$

  The parameter a for the projected marginal diameter distribution at time t will be half of $\hat{D}_{min,t}$.

- Predict the parameter c for the marginal three-parameter Weibull Distribution using the estimated quantile $\hat{q}_{25,t}$ and $\hat{q}_{95,t}$ with the following equation

$$\hat{c} = \frac{ln\left(\frac{ln(1-0.25)}{ln(1-0.95)}\right)}{(ln(\hat{q}_{25,t} - \hat{D}_{min,t})) - (ln(\hat{q}_{95,t} - \hat{D}_{min,t}))} \quad (5.27)$$

- With $\hat{D}_{min,t}$ and $\hat{c}$, predict the parameter b using the following equation.

$$\hat{b} = -\left(\frac{\hat{D}_{min,t}\Gamma_1}{\Gamma_2}\right) + \sqrt{\left(\frac{\hat{D}_{min,t}}{\Gamma_2}\right)^2 (\Gamma_1^2 - \Gamma_2) + \left(\frac{Dq_t^2}{\Gamma_2}\right)} \quad (5.28)$$

  where $\Gamma_k = (1 + \frac{k}{\hat{c}})$, $k = 1, 2$.

- Once the two Weibull distributions for t_0 and t are estimated parametrically, we can define the continuous parametric copula C^s with known margins. Let's define $(\boldsymbol{\beta}, \theta)$ as the set of all the parameters required to define the marginal Weibull at time t and the copula. We can found the likelihood of the parameters under the model assumption given the data. The log-likelihood is represented as

$$\text{log-lik}_n(\boldsymbol{\beta}, \theta|\text{data}) = \sum_{i=1}^{n} ln\acute{c}_\theta(F_1(X_{i,t_0}), F_{2,\boldsymbol{\beta},t}) + \sum_{i=1}^{n} lnf_1(X_{i,t_0}) + \sum_{i=1}^{n} lnf_{2,\boldsymbol{\beta},t}(X_{i,t}). \tag{5.29}$$

where $\acute{c}_\theta$ is the density of C^s, and $f_{j,\boldsymbol{\theta}_j}$ is the density of F_j.

- The parameter vector $(\boldsymbol{\beta}, \theta)$ can then be estimated by the Maximum likelihood estimator (MLE), that is, the parameters that make maximum the equation in (5.29).

- The marginal Weibull for SDD at time t can then be projected from the any initial SDD at time t_0 using the estimated parameters and equations (5.26), (5.27), and (5.28).

The proposed method allows finding a parametric continuous bi-dimensional copula with Weibull marginals. Once the copula's parameter(s) and one of its marginal are known, it is possible to simulate diameter values for the other marginal distribution or recover the parameters using projected stand information and the estimate copula's parameter. Interestingly, it is possible then to use the copula's parameter(s) to project the Weibull marginal survival diameter distribution given the actual surviving trees' information.

5.2.4 Validation

A subset of 14 first-thinning installations from the MRT study was randomly selected as the fit data set. We fitted the parameters in equations (5.23), (5.24), (5.25), and (5.26) using the fit data set. All the possible combinations of pairs of measurements within the plot (of type (t_i, t_j) with $i < j$) were used for parameter estimation. A validation data set consisted of the remaining 11 first-thinning installations and all their possible combinations of pairs of measurements within the plot.

The method presented above was evaluated against the observed diameters at age t, assuming the observed diameter distribution was given at age t_0. The traditional method

of Parameter Recovery (PR) was implemented to compare the accuracy of this new method with copulas. We implement the PR method proposed by (Qin et al., 2007) using the fit dataset.

Two forms of error-index proposed by (Reynolds et al., 1988) were used to determine how well the two methods (Distribution projection with copulas and the traditional PR) perform when predicting information at the dbh-class level. The error-index in terms of number of trees per hectare (e_TPH_i) for the ith plot is

$$e\_TPH_i = \sum_{j=1}^{k_i} |TPH_{ij} - \hat{TPH}_{ij}| \tag{5.30}$$

The error-index in terms of basal area per hectare (e_BA_i) for the ith plot is

$$e\_BA_i = \sum_{j=1}^{k_i} |BA_{ij} - \hat{BA}_{ij}| \tag{5.31}$$

where TPH_{ij} and $\hat{TPH}_ij$ are the observed and predicted number of trees per hectare of the jth diameter class in the ith plot, respectively. BA_{ij} and $\hat{BA}_{ij}$ are the observed and predicted stand basal area of the jth diameter class in the ith plot, respectively. Each diameter class's width was fixed at 3 cm, k_i represents the number of 3 cm diameter classes in the ith plot.

5.3 Results

5.3.1 Copula Selection

The copula selection analysis included all the plot measurements of the MRT study. We were particularly interested in testing if the underlying bivariate dependence structure could change by physiographic region or treatment type. Table 5.1 summarizes the result for the Exchangeability, Symmetry, and Extreme-Value tests, for all the observed within-plot pairs

SDD (t_0, t). The counts correspond to number of plots with $p - value < \frac{\alpha}{m}$, where m is the number of comparisons in each group.

Table 5.1: Count of tests for Exchangeability, Symmetry, and Extreme-Value properties, where the null hypothesis is not plausible (numerator) of the total tests performed (denominator). T1 is the control for first thinning condition, T2 = thinning only, T3 = thinning + fertilization, T4 = thinning + CVC, T5 = thinning + fertilization + CVC; LC = Lower Coastal Plain, PI = Piedmont, and UC = Upper Coastal Plain.

	Treatment				
Region	**T1**	**T2**	**T3**	**T4**	**T5**
	Exchangeability test				
LC	0/24	0/24	0/24	0/24	0/24
PI	0/9	0/9	0/9	0/9	0/11
UCP	0/24	0/24	0/24	0/24	0/24
	Symmetry test				
LC	0/24	0/24	0/24	1/24	0/24
PI	0/9	3/9	0/9	0/9	0/11
UCP	0/24	0/24	0/24	1/24	0/24
	Extreme-Value test				
LC	22/24	2/24	0/24	0/24	1/24
PI	8/9	1/9	0/9	0/9	0/11
UCP	22/24	0/24	0/24	0/24	0/24

Analyzing the preliminary exploratory analysis on scatter plots between the transformed survival diameter observations (pseudo-observations) in each pair of samples between time t_0 and times t, like Figure (5.1), demostrates the presence of strong diameter correlations between measurements and the prevalence of symmetric tails. Results in Table 5.1 confirm that it is extremely likely that the underlying unknown survival diameter copulas in the MRT for all treatments are exchangeable and symmetric.

Extreme-Value test results indicated that some evidence against the null hypothesis about Extreme-Value dependence in the control plots could be considered. Although it does not say anything about the strength of the Non-Extreme-Value dependence, the result suggests that the hypothesis of Extreme-Value dependence could not be considered for all the non-thinned plots. That is, 89% to 92% of tests within non-thinned plots showed evidence against Extreme-Value dependence. In contrast, at most 0.5% of the tests within thinned plots

showed evidence against Extreme-Value dependence (see Table 5.1). The following results about goodness-of-fit will help to decide if exist some change in the type of dependence between the thinned and non-thinned treatments and how the chosen copulas fit the data.

Table 5.2 summarizes the goodness-of-fit tests (Equation 5.22) using five candidate copulas to model the survival diameter dependence. The result is the count of tests that returned a $p - value$ lower than α/m, where $\alpha = 0.05$ and m is the number of comparisons in each combination physiographic region and treatment type.

Table 5.2: Count of the goodness-of-fit tests where the null hypothesis (the observed dependence belongs to one of the copula family) is not plausible (numerator) of the total tests performed (denominator). T1 is the control for first thinning condition, T2 = Thinning only, T3 = Thinning + Fertilization, T4 = Thinning + CVC, T5 = Thinning + Fertilization + CVC; LC = Lower Coastal Plain, PI = Piedmont, and UC = Upper Coastal Plain.

	Treatment					**Treatment**				
Region	**T1**	**T2**	**T3**	**T4**	**T5**	**T1**	**T2**	**T3**	**T4**	**T5**
	Clayton copula					**Gumbel-Hougaard copula**				
LC	24/24	16/24	21/24	19/24	15/24	7/24	0/24	0/24	2/24	0/24
PI	9/9	9/9	9/9	8/9	9/11	6/9	0/9	0/9	0/9	0/11
UCP	23/24	21/24	19/24	17/24	16/24	13/24	0/24	0/24	0/24	0/24
	Normal copula					**t copula**				
LC	3/24	0/24	0/24	2/24	0/24	0/24	0/24	0/24	1/24	0/24
PI	0/9	4/9	0/9	0/9	0/11	0/9	3/9	0/9	0/9	0/11
UCP	4/24	1/24	0/24	0/24	0/24	1/24	0/24	0/24	1/24	0/24
	Galambos copula									
LC	6/24	0/24	0/24	1/24	0/24					
PI	4/9	0/9	0/9	0/9	0/11					
UCP	13/24	0/24	0/24	0/24	1/24					

A significant proportion of tests for the copula family class Clayton indicates that the data are not compatible with this survival diameter dependence. Clayton's family is characterized by having an upper tail high dependence and less dependence in the lower tail. However, the tail dependence in survival diameter looks symmetric and highly dependent on both tails, which might explain the largest number of cases where there is strong evidence against Clayton's dependence on the studied plots. The Gumbel-Hougaard copula family fit better the survival diameter dependence for thinned treatments plots than for non-thinned control

plots. The two elliptical families (Normal and Student t) show a better fit for the survival diameter dependence. The t copula hypothesis is not supported in 6 of 287 tests, and the Normal copula is not supported in 14 of 287 tests (see Table 5.2). The Extreme-value copula Galambos behave similarly to the Gumbel-Hougaard copula. For treated thinned plots, the Extreme-value hypothesis shows an acceptable fit. However, for some non-thinned plots, the hypothesis of this type of dependence is not plausible. The lack of a good fit for the Galambos copula in non-thinned control plots is consistent with the Extreme-value dependence tests results in Table 5.1. We did not identify any pattern of lack of goodness-of-fit across the regions.

We concentrate our copula selection on families Gumbel, Normal, t, and Galambos. Table 5.3 shows the results for leave-one-out cross-validation CIC by treatment. For treatments T1 (control plots), the t copula ranking best. However, in the presence of other mid-rotation treatment (Fertilization in T3, CVC in T4, or combination in T5), the Normal and Gumbel perform similar or best than the t copula. The Galambos copula does not perform well with treatments T1 and T5. Previous results showed in Chapter 4 indicated a positive response in growth due to the treatments compared to thinning only on this same dataset. This might explain the change in family dependence, given the presence of MRT. Considering the two tests above, we can conclude that the Normal copula could be adequate to model the dependence in treated plots, and the Student t copula could be used for control and thinning only plots.

The Student t copula has two parameters, the correlation coefficient ρ, that is in the off-diagonal of a correlation matrix, and the $v > 0$ degrees of freedom. So far, we do not find a logical and meaningful relationship between our SDD projection theory and the parameter v of the multivariate t copula. Therefore, in this study, we selected the Normal Copula to model the distribution dependence.

Table 5.3: Results of Copula Information Criteria (CIC) by treatment (rows) with four family copulas (Gumbel, Normal, Studnet t, and Galambos). The values represents the number of times that a copula family produces the largest CIC. T1 is the control for first thinning condition, T2 = Thinning only , T3 = Thinning + Fertilization, T4 = Thinning + CVC, T5 = Thinning + Fertilization + CVC.

	Family				
Treatment	**Gumbel**	**Normal**	**Student t**	**Galambos**	**Total fitted copulas**
T1	1	10	46	0	57
T2	19	5	18	15	57
T3	9	16	14	18	57
T4	20	15	8	14	57
T5	18	24	15	2	59

5.3.2 Diameter Distribution Projection

The fit dataset consisted of 295 plot-measurement periods from 70 plots in 14 installations. The validation dataset consisted of the remaining 11 installations, with 217 plot-measurements from 57 plots.

The system of four equations, (5.23), (5.24), (5.25), and (5.26) were fitted simultaneously. The algorithm "L-BFGS-B" from the R package `optimParallel` (Gerber and Furrer, 2019) was used to find the parameters that minimize the log-likelihood for the Normal Copula over all the possible pairs periods in the fit dataset.

Table 5.4 presents the estimated parameters for equations (5.23), (5.24), (5.25), and (5.26). The respective standard errors were found using bootstrap methods as not all the parameters were completely identifiable with the inverse of the Hessian matrix. One thousand random replications of the complete fit data set were obtained sampling with replacement, then the model system was fitted with each replication, and the fitted parameters were stored. The approximated standard error was obtained as:

$$SE_{Boot}(\theta_j) = \sqrt{\frac{\sum_{i=1}^{N}(\theta_{ij}^{*} - \hat{\theta}_j)^2}{N}} \tag{5.32}$$

where, $SE_{Boot}(\theta_j)$ is the bootstrap standard error for the jth parameter, θ^*_{ij} is the estimate jth parameter from the ith replicate data set, $\hat{\theta}_j$ is the estimate parameter with the original data set, N is the number of replications.

Table 5.4: Parameter estimated and standard errors (SE) for the system of equations in equations (5.23), (5.24), (5.24), and (5.26).

Parameter	**Estimate**	**SE**	**Parameter**	**Estimate**	**SE**
$\hat{\beta}_1$	0.03931	5.4556e-03	$\hat{\beta}_8$	1.05552	1.8805e-02
$\hat{\beta}_2$	-0.19145	3.2423e-02	$\hat{\beta}_9$	0.97727	1.4510e-02
$\hat{\beta}_3$	0.01073	1.3078e-03	$\hat{\beta}_{10}$	-0.69444	1.6819e-02
$\hat{\beta}_4$	0.25411	1.0133e-02	$\hat{\beta}_{11}$	0.64221	1.4661e-02
$\hat{\beta}_5$	1.23324	2.4110e-02	$\hat{\beta}_{12}$	0.23322	4.9756e-02
$\hat{\beta}_6$	-1.54007	8.7922e-03	$\hat{\beta}_{13}$	1.62102	6.9941e-02
$\hat{\beta}_7$	1.14855	7.3656e-03	$\hat{\beta}_{14}$	-1.40819	4.6797e-02

The results so far indicate that all the independent variables are important in the models. Residuals analysis (no-shown) does not indicate anomalies concerning the residual variance. Therefore no additional weight variance function was included in the fitting process.

5.3.3 Validation

Table 5.5 presents the mean, standard errors, and 95% confidence intervals of Reynolds et al.'s (1989) error-index values from the validation dataset using the method of diameter projections with copula and PR.

The method of parameter recovery (PR) and the proposed method using copulas produce similar error-index values for trees per hectare and basal area when summarized by treatment based on comparison of their confidence intervals (see Table 5.5). Improvements on standard deviations for the mean with the copula method indicate a gain in accuracy on predictions and consequently narrow confidence intervals for error-index. The two methods are consistent throughout the treatments, which indicate that both are well suited to use with plantations managed with mid-rotations treatments.

Table 5.5: Reynolds et al.'s (1989) error-index for trees per hectare ($e-TPH$) and basal area per hectare ($e-BA$) from two methods and by treatment for the MRT study validation data set. **n** amount of pairs observations for validation, **sd** standard deviation.

Treat.	**Method**	**n**	$e-TPH$ **(trees ha^{-1})**	**sd**	**95% CI**	$e-BA$ **(m^2 ha^{-1})**	**sd**	**95% CI**
T1	PR	43	211	77.2	(187, 235)	7.5	2.8	(6.6, 8.4)
	Copula	43	201	69.0	(180, 222)	7.4	3.0	(6.5, 8.3)
T2	PR	43	145	62.5	(126, 164)	7.5	2.9	(6.6, 8.4)
	Copula	43	103	40.6	(91, 115)	4.9	1.6	(4.4, 5.4)
T3	PR	43	129	52.8	(113, 145)	7.0	3.0	(6.1, 7.9)
	Copula	43	126	46.8	(112, 140)	6.5	2.4	(5.8, 7.2)
T4	PR	43	131	49.3	(116, 146)	7.2	2.9	(6.3, 8.1)
	Copula	43	104	38.2	(92, 116)	5.2	1.9	(4.6, 5.8)
T5	PR	45	107	50.7	(92, 122)	6.0	2.4	(5.3, 6.7)
	Copula	45	95	38.3	(83,107)	5.1	1.9	(4.5, 5.7)

Figure 5.2 shows the error-index $e-TPH$ and $e-BA$ values grouped by projection time. As the projection time increases, the error-index increases as well, especially for basal area predictions. Although there are fewer measurements at the most prolonged time intervals, the copula method appear to be more robust. That is, produce a lower error-index compared with PR as the projection time increases.

5.3.4 Application: Predicting Volume by Products

Foresters and biometricians are also interested in assessing future product values in the managed plantations. Chapter 4 presented a system of equations to project the dominant height, number of trees per hectare, and total stand basal area after treatment. The presented method of Diameter Distribution projection using copulas can represent the overall stand level trees per hectare and basal area by dbh-classes. The taper equation presented in Chapter 3 can be used to find the total and merchantable volume for the average tree in each dbh-class.

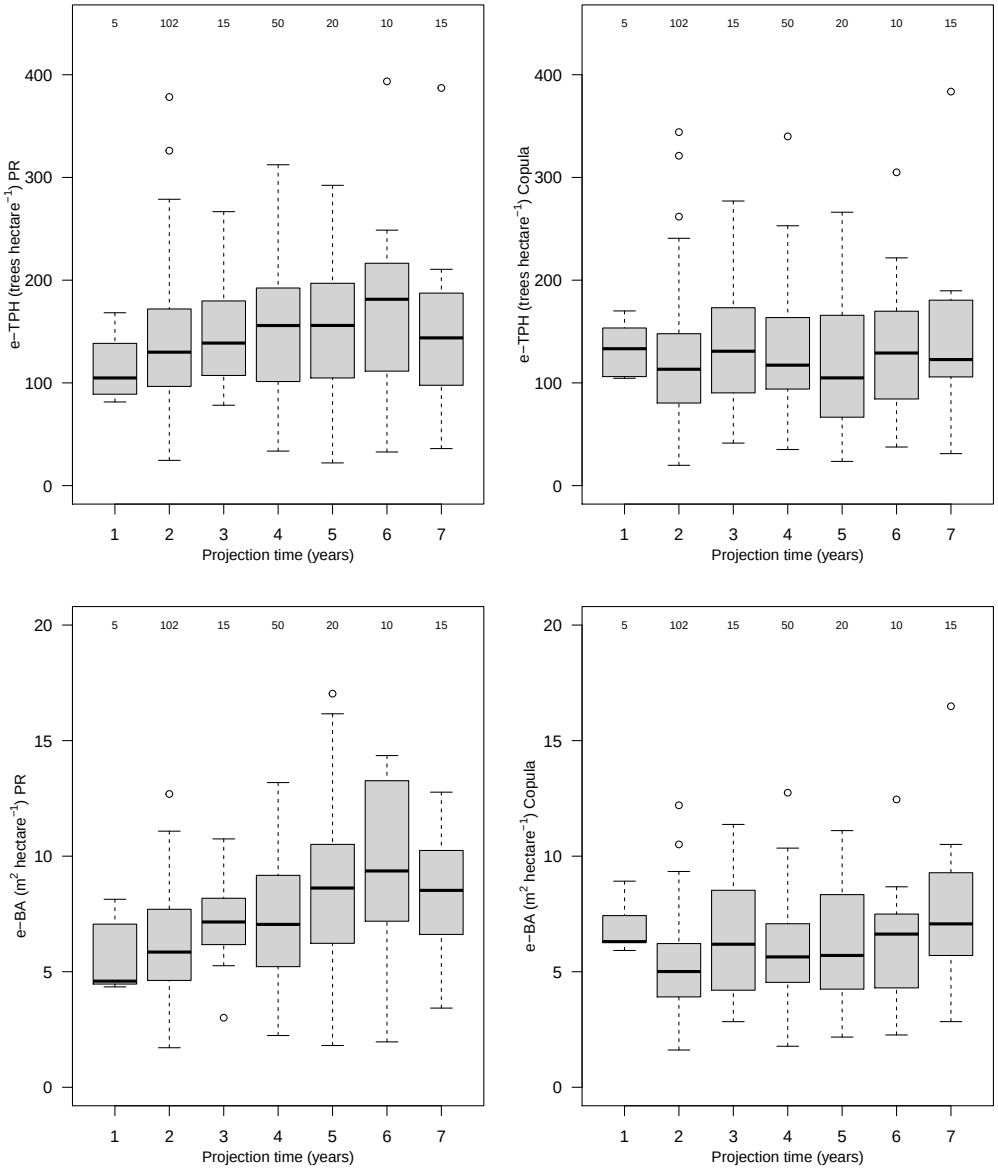

Figure 5.2: Reynolds et al.'s (1988) error-indices based on number of trees per hectare ($e-TPH$)(top row) and basal area per hectare ($e-BA$) (bottom row) at different projection times with the validation dataset. The plots on the left correspond to the Parameter Recovery (PR) method and the plots on the right with the copula method. The numbers on the top of each box are the available observations for validation in each projected time.

Let us assume that a loblolly pine stand located in the southeastern U.S. in the physiography region Lower Coastal Plain received a mid-rotation thinning treatment at age 15. Table 5.6 shows the known stand-level information before and after the treatment.

Table 5.6: Stand level information for a hypothetical loblolly pine plantation in the Lower Coastal Plain southeastern of the U.S. receiving a thinning treatment at age 15.

Stand attribute	**Before Thinning**	**After Thinning**
Age	15	15
Trees per hectare	1493 trees	489 trees
Basal Area	26.7 m^2 ha^{-1}	11.5 m^2 ha^{-1}
DH	15.5 m	15.5 m
R		0.672
Dq	15.09 cm	17.30 cm
D minimum	7.62 cm	15.49 cm
Weibull distribution		
a	3.81	7.747
c	3.875	5.93
b	14.860	13.04
Q_{10}	12.12	16.67
Q_{25}	14.58	18.32
Q_{95}	23.53	23.43

We are interested in projecting the stand level information to age 25 (rotation age) to evaluate the availability of total wood volume and merchantable volume, considering the five treatments presented in the MRT study. The wood products were defined as

- Pulpwood: 15+ cm DBH to a 7 cm i.b. top
- Chip-n-saw: 23-29 cm DBH to a 15 cm i.b. top
- Sawtimber: 30+ cm DBH to a 20 cm i.b. top

Table 5.7 shows the projected attributes at age 25, including the estimated parameters for the projected Weibull marginal diameter distribution by treatment.

Table 5.8 shows the total volume and merchantable volume per hectare at age 25, resulting from the stand diameter distribution and classifying the diameters classes according to

Table 5.7: Projected stand-level information for a hypothetical loblolly pine plantation in the Lower Coastal Plain southeastern U.S. at age 25 for five different treatments (T1= No treatment, T2= Thinning only, T3 = Thinning and Fertilization, T4 = Thinning and CVC, T5 = combination of Thinning, Fertilization and CVC).

Treatment	**BA** ($\mathbf{m^2\ ha^{-1}}$)	**DH** (m)	**TPH**	**Dq** (cm)	**Q25** (cm)	**Q95** (cm)	$\hat{a}$	$\hat{c}$	$\hat{b}$
T1	39.74	25.4	1009	22.4	19.5	31.2	5.06	3.83	17.97
T2	25.29	25.87	472	26.1	24.9	33.5	10.99	4.88	16.24
T3	27.95	25.87	472	27.5	25.5	34.2	10.99	5.03	17.66
T4	27.64	25.87	472	27.3	25.5	34.1	10.99	5.01	17.49
T5	28.80	25.87	472	27.9	25.7	34.4	10.99	5.08	18.09

Note: The projected basal area, dominant height, and trees per hectare were found using the system of growth and projection equations presented in Chapter 4; $\hat{a}$, $\hat{c}$, and $\hat{b}$ are the estimated three-parameter of marginal Weibull distribution obtained with equations (5.26), (5.27), and (5.28) respectively.

the DBH product's limits. The results indicate that thinning and the combination of thinning with other silvicultural treatment increases the proportion of total volume allocated in merchantable volume other than pulp. Thinning + Fertilization (T3) and Thinning + CVC produce similar proportional results allocating total volume in wood products. The combinations of treatments (T5) favors individual tree growth and produces the largest accumulation in Sawtimber (17.3% more than thinning only; see Table 5.8).

This simulated exercise clearly represents these silvicultural treatment objectives, increasing growth, and adding value to the plantation through management. For example, thinning combinations with fertilization or CVC allocate between 91 to 93% of total volume in products Chip-n-saw and Sawtimber combined. Thinning alone allocates 87.1% of the total volume in those two categories, while not performing any mid-rotation treatment allocates less volume in these two products, 61.2% of the total. Optimal management decisions depend on management objectives, management costs, location, availability of resources, and market conditions. Here we offer a tool that allows for a more accurate projection of merchantable volume, which is essential for decision making.

Table 5.8: Total and merchantable volume per hectare at age 25 from a hypothetical loblolly pine plantation located in Lower Coastal Plain. Results for five mid-rotation treatments. (T1= No treatment, T2= Thinning only, T3 = Thinning and Fertilization, T4 = Thinning and CVC, T5 = combination of Thinning, Fertilization, and CVC). In parenthesis is the percent of the value with respect to the total volume per row.

Treatment	**Total Volume** (m^3 ha^{-1})	**Pulp** (m^3 ha^{-1})	**Chip-n-saw** (m^3 ha^{-1})	**Sawtimber** (m^3 h^{-1})
T1	377.71	139.13 (36.8%)	178.33 (47.2%)	52.75 (14.0%)
T2	248.82	32.02 (12.9%)	189.73 (76.3%)	26.81 (10.8%)
T3	277.04	22.11 (8%)	189.85 (68.5%)	64.85 (23.4%)
T4	274.20	23.04 (8.4%)	191.45 (68.8%)	59.46 (21.7%)
T5	287.72	19.95 (6.9%)	186.68 (64.9%)	80.86 (28.10%)

Note: Total Volume is the volume inside bark; The DBH size-classes were 3cm wide product classes were determined on ly based on DBH; The middle tree volume in each DBH size-class was obtained with the semiparametric taper equation presented in Chapter 3; The following equation was fit with all the tree height measurements from the MRT dataset and was used to predict the middle size-class tree height: $h = 1.068365 * DH * [1 - 1.485925 * exp(-2.801092 * (DBH/Dq) - 0.193472 * IT)]$, DH is the stand dominant height, Dq is the stand mean quadratic diameter, IT is an indicator variable, $IT = 1$ if the stand was thinned, $IT = 0$ for non thinned stands, RMSE=0.9653, and $n = 24362$.

5.4 Discussion

Several methods have been proposed to model diameter distributions in forest stands. Fitting a probability density function to the observed diameter structure is the most accepted method. After that Bailey (1973) proposed for the first time using the three-parameter Weibull distribution as an adequate probability density function for diameter distributions, it has been extensively used in forest modeling, and alternative parameter prediction methods have been tested.

Methods like parameter recovery or percentile prediction procedures are the most used because they allow relating stand characteristics with pdf statistics like moments of the distribution or specific quantiles. Typically, these relationships are based on regression-like methods, therefore expected pdf for future diameter distribution are predicted from initial and projected stand characteristics using stand-level models. However, the mentioned meth-

ods ignore the current distribution information (if measured) and dependence information are not directly used to predict the future distribution (Knoebel and Burkhart, 1991).

We suggest using a copula to model the dependence between observed survival trees at two points in time. The copula is a method used in statistics to represent p dimensional process and study the dependence between p dimensions. Knoebel and Burkhart (1991) proved a similar modeling strategy, but they used bivariate Johnson's S_B (Johnson, 1949). The bivariate S_B has the property that the conditional distribution of diameters given the initial distribution is also S_B. However, copulas are more versatile, and we can combine different continuous *pdf* as marginals, i.e, Gamma, Normal, or Weibull (studied here).

The study of the dependence of distribution over time is both intuitive and biologically reasonable. The future shape and location of the SDD are highly related to the type and intensity of plantation management. Projections over short periods should result in higher dependence compared with long projection periods. We used the MRT first-thinning dataset to evaluate the SDD dependence through time. We found that the copulas for SDD are symmetric and exchangeable. This is explained in part because each observation in this framework is a pair of measured diameters from the same tree. Finally, the study of dependence was for surviving trees; it means that the resulting diameter projections have the effect of mortality implicitly.

The proposed method for surviving diameter distribution projections was based on Normal copulas. Student t copulas also works well for SDD, but future research is needed to relate the degrees of freedom with stand-level characteristics. The validation with Normal copulas suggests that the proposed method and the traditional parameter recovery method provide equally good fits to the observed trees per hectare and basal area by type of treatment. However, projections with the copula method are more consistent. They produce fewer projection errors when the projection is far from the initial observed distribution. Fo-

erst modelers shall find this result more convenient to assess future wood supply accurately for decision-making.

5.5 Conclusions

In this research, a new approach to project diameter distribution of future stands was presented. We used copulas to model the dependence of surviving trees at two-points in time, and then the copula parameter was associated with stand-level attributes. A set of equations for recovery of one of the marginal distribution for the future stand was presented. The copula method produces similar and sometimes better results than the traditional parameter recovery method in terms of Reynolds et al.'s (1988) error-index values. When comparing plots with different mid-rotation treatments, notable improvement in accuracy was identified with the copula's method as the projection period increases (maximum of 7 years in this study). The main improvement over current methods is that the initial state's distribution information is not considered independent of the method used to project future stand distribution, and the assumed dependence is considered to recover the parameters of the future marginal diameter distributions.

References

Álvarez González, J., Schröder, J., Rodriguez Soalleiro, R., and Ruiz González, A. (2002). Modelling the effect of thinnings on the diameter distribution of even-aged maritime pine stands. *Forest Ecology and Management*, 165(1):57–65.

Bailey, R. L. (1973). Quantifying diameter distributions with the weibul funciton. *Forest Science*, 19(2):97–104.

Ben Ghorbal, N., Genest, C., and Neslehova, J. (2009). On the Ghoudi, Khoudraji, and Rivest test for extreme-value dependence. *Canadian Journal of Statistics*, 37(4):534–552.

Binoti, D., Marques da Silva, M., Garcia, H., Fardin, L., and Castro, J. (2012). Probability density functions for description of diameter distribution in thinned stands of Tectona grandis. *Cerne, Lavras*, 18(2):185–196.

Breidenbach, J., Gläser, C., and Schmidt, M. (2008). Estimation of diameter distributions by means of airborne laser scanner data. *Canadian Journal of Forest Research*, 38(6):1611–1620.

Bullock, B. P. and Burkhart, H. E. (2005). Juvenile diameter distributions of loblolly pine characterized by the two-parameter weibull function. *New Forests*, 29(3):233–244.

Burkhart, H. E. and Tomé, M. (2012). *Modeling forest trees and stands*. Springer. OCLC: ocn799874477.

Cao, Q. V. (2004). Predicting parameters of a weibull function for modeling diameter distribution. *Forest Science*, 50(5):682–685.

Cao, Q. V. (2006). New method for estimating parameters of weibull function to characterize future diameter distributions in forest stands. In *Connor, K. F. ed:*, SRS-92, page 640. Forest Service, Southern Research Station.

Cohen, A. (1965). Maximum likelihood estimation in the weibull distribution based on complete and on censored samples. *Technometrics*, 7(4):579–588.

Duan, A.-g., Zhang, J.-g., Zhang, X.-q., and He, C.-y. (2013). Stand diameter distribution modelling and prediction based on richards function. *PLOS ONE*, 8(4):e62605.

Frazier, J. R. (1981). *Compatible whole-stand and diameter distribution models for loblolly pine plantations*. Doctor of Philosophy in Forestry, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Genest, C., Nešlehová, J., and Quessy, J.-F. (2012). Tests of symmetry for bivariate copulas. *Annals of the Institute of Statistical Mathematics*, 64(4):811–834.

Genest, C. and Nešlehová, J. G. (2014). On tests of radial symmetry for bivariate copulas. *Statistical Papers*, 55(4):1107–1119.

Genest, C., Rémillard, B., and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, 44(2):199–213.

Gerber, F. and Furrer, R. (2019). optimparallel: An r package providing a parallel version of the l-bfgs-b optimization method. *The R Journal*, 11(1):352–358.

Grønneberg, S. and Hjort, N. L. (2014). The copula information criteria. *Scandinavian Journal of Statistics*, 41(2):436–459.

Harter, H. and Moore, A. (1965). Maximum-likelihood estimation of the parameters of gamma and weibull populations from complete and from censored samples. *Technometrics*, 7(4):639–643.

Hofert, M., Kojadinovic, I., Mächler, M., and Yan, J. (2018). *Elements of copula modeling with R*. Use R! Springer International Publishing, Cham.

Johns, M. and Lieberman, G. (1966). An exact asymptotically efficient confidence bound for reliability in the case of the Weibull distribution. *Technometrics*, 8(1):135–175.

Johnson, N. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, 36(1):149–176.

Knoebel, B. R. and Burkhart, H. E. (1991). A Bivariate distribution approach to modeling forest diameter distributions at two points in time. *Biometrics*, 47(1):241–253.

Luko, S. N. (1999). A review of the weibull distribution and selected engineering applications. Technical Report 1999-01-2859, Society of Automotive Engineers, Inc., Indianapolis, Indiana.

Marius Hofert and Martin Mächler (2011). Nested archimedean copulas meet R: The nacopula package. *Journal of Statistical Software*, 39(9):1–20.

Mateus, A. and Tomé, M. (2011). Modelling the diameter distribution of eucalyptus plantations with johnson's SB probability density function: parameters recovery from a compatible system of equations to predict stand variables. *Annals of Forest Science*, 68:325–335.

Mctague, J. P. and Bailey, R. L. (1987). Compatible basal area and diameter distribution models for thinned loblolly pine plantations in santa catarina, Brazil. *Forest Science*, 33(1):43–51.

Owzar, K. and Sen, P. K. (2013). Copulas: concepts and novel applications. *METRON-International Journal of Statistics*, LXI(3):323–353.

Qin, J., Cao, Q. V., and Blouin, D. C. (2007). Projection of a diameter distribution through time. *Canadian Journal of Forest Research*, 37(1):188–194.

Reynolds, M., Burk, T. E., and Hung, W.-C. (1988). Goodness-of-fit test and model selection procedures for diameter distribution models. *Forest Science*, 34(2):373–399.

Schreuder, T. H. and Hafley, L. W. (1977). A useful bivariate distribution for describing stand structure of tree heights and diameters. *Biometrics*, 33(3):471–478.

Weibull, W. (1939). A statistical theory of the strength of materials. Technical Report 151, Ing. Vetensk. Akad. Handl., Sweden.

Chapter 6

Conclusions

It is well known that mid-rotation silvicultural plantation management like vegetation competition control, fertilization, and thinning impact the expected future stand growth and yield. Therefore, it is crucial to understand how to incorporate these responses and their interactions inside the G&Y predictions models.

The G&Y system presented here consists of a novel taper equation based on a penalized spline regression, a compatible dynamic growth system of differential equations for dominant height, basal area, and stand density, including a growth modifier to account for mid-rotation silvicultural effects, and a novel method to recover the diameter distribution for projected stand after mid-rotation silvicultural treatments.

A semiparametric approach was proposed at the tree level to model the tree stem form of loblolly pine. The model is a penalized spline regression, P-Spline, of degree $p = 3$. We also tested the performance of an extended P-Spline with an additive dbh-class factor variable. The proposed extended P-Spline method outperforms the traditional parametric taper equations when used to predict outside bark diameters in the lower portions of the stem, up to 40% of the tree relative height where the more valuable wood products (on average 62% of the total outside bark volume) are contained.

While the inclusion of a dbh-class variable improves the P-Spline fitting and their capability to explain taper shapes, this comparative advantage was not reflected in a more accurate ability in terms of volume predictions. Instead, the simple P-Spline showed the best performance in terms of volume. However, the proposed extended P-Spline performs similarly to the best parametric equation by Max and Burkhart (1976) with taper calibration.

Although taper calibration with extra upper diameter measurements has proved to increase volume prediction accuracy, the simple P-Spline model presented here shows superior performance without any additional calibration. Consequently, the use of semiparametric taper modeling could result in savings on inventory costs by omitting any additional measurements. Additionally, the *a priori* fixed forms assumed by the empirical taper equations imposes an unnecessary restriction that fails to explain the tree form adequately compared with P-Spline. Further, we expect that the use of P-Spline for loblolly pine volume estimations will be more reliable because it minimizes error due to model misspecifications.

The modeling strategies followed in this research at the stand-level offer a biological and mathematical consistent framework to assess loblolly pine growth responses due to mid-rotation treatments, including thinning, fertilization, competitive vegetation control, and their combinations. The proposed model system is a dynamic compatible growth and yield system rooted in theoretical and biologically principles. The model proved highly predictive accuracy and can be used to project the current stand attributes following any combinations of mid-rotation silvicultural practices with different thinning levels and intensities. The modifier incorporated in the growth model for the basal area includes a temporal effect of treatments. Because of the model structure, the responses to treatments change with location, age when applied, and dominant height growth.

The projected whole-stand structure obtained with the previous G&Y system needs to be described by a diameter distribution to produce low-level details and determine wood product composition and value. Rather than independently predicting the future diameter

distribution, this dissertation suggests using a copula to model the dependence between observed survival diameter distributions of trees at two points in time. The study of the dependence of distribution over time is both intuitive and biologically reasonable.

The copula method produces similar and sometimes better results than the traditional parameter recovery method in terms of Reynolds et al.'s (1988) error-index values. When comparing plots with different mid-rotation treatments, notable improvement in accuracy was noticed with the copula's method as the projection period increase (maximum of 7 years in this study). The main improvement over the current methods is that the initial state's distribution information is not considered independent of the method used to project future stand distribution, and the assumed dependence is used to recover the parameters of the future marginal diameter distributions.

Overall the new models presented in this dissertation contribute to the understanding of plantation growth dynamics managed with mid-rotation silvicultural treatments in the southeastern of the U.S.

References

Max, T. and Burkhart, H. (1976). Segmented polynomial regression applied to taper equations. *Forest Science.*, 22(3):283–289.