Mortality Modeling in Douglas-fir Plantations in the Pacific Northwest

Bridgette Hawes

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

University of Washington

2021

Committee Members:

Chair: Eric Turnblom

Gregory Ettl

Patrick Clune

Program Authorized to Offer Degree:

School of Environmental and Forest Sciences

University of Washington

Abstract

Mortality Modeling in Douglas-fir Plantations in the Pacific Northwest

(Pseudotsuga menziesii var. menziesii)

Bridgette Alexandra Hawes

Chair of Supervisory Committee:

Dr. Eric Turnblom

School of Environmental and Forest Sciences,

College of the Environment, University of Washington

Forest management is dynamic and complex, with numerous factors affecting forest development and outcomes of management decisions. In plantation forestry, forest development is typically expected to follow well understood patterns. However, these patterns may vary substantially due to unpredictable factors. One such factor that is difficult to accurately anticipate is tree mortality. To facilitate the ability of managers to make the best management decisions, mortality must be effectively estimated over the life of the stand. This is often done using forest growth and yield projections, however most growth and yield modeling techniques oversimplify the process of accounting for stand-level mortality. The objective of this project was to create a deterministic mortality model for Douglas-fir

(*Pseudotsuga menziesii [Mirb.] Franco*) plantations in the Pacific Northwest that can differentiate the pattern of mortality from the starting planting density. Mortality models have existed for decades, all taking different approaches to modeling and validation. Multiple new mortality modeling techniques were explored throughout this study, using non-linear sigmoidal curves to capture the trend of mortality over time. Using a parameter prediction approach and validating the process with a bootstrapping procedure, a final model was selected based on the R^2, MAE and SD values. Both absolute and relative models were used in the overall process before determining that an absolute model was capturing the behavior of mortality better over time and moved forward with. The final model will be coupled with the Stand Management Cooperatives Plantation Yield Calculator--a growth and yield model-- to improve overall estimates to aid in management decisions in Douglas-fir stands.

Contents:

List Of Figures

List of Tables

Chapter 1. Introduction:

The density at which forest managers plant a stand is crucial in determining the subsequent silvicultural treatments that will be applied throughout the rotation. Planting at a low initial density can limit your management options over a course of a harvest rotation, as well as cause intense early competition with vegetation in the early initiation stage of stand development and can have substantial effects on merchantability of the timber harvested (Harrington et al. 2009). A high initial planting density allows the trees planted to dominate the growing space, limiting competing vegetation. A high initial planting density also controls for early growth form which can lead to more merchantable trees. Though this may also cause stand-level growth rates to slow quicker and faster in comparison to a lower density stand and will create the need for a pre-commercial thin in order to re-open the growing space back up before the stand-level growth starts to decline. Understanding these trade-offs, determining what density to plant at is a crucial management decision. However, mortality of trees significantly complicates planting density decisions. Managers expect some death of the seedlings due to natural causes (browse, competition, sun scorch, drought) but the factors driving mortality have yet to be fully quantified into a predictive model over its in entire rotation. There are many mortality models for different ecosystems which look at different potential influences of mortality (e.g., wind, fire, insects, droughts; Temperli et al. 2013, Seidl et al. 2017, Bigler et al. 2006). These include tree and stand-level mortality models, as well as mortality models focused on sources of mortality such as diseases and pest outbreaks. The goal was to create a model using long-term data from planting to rotation age to be able to accurately account for Douglas-fir mortality in the maritime Pacific Northwest over an entire rotation. This is not an individual tree mortality model like many have done in the past, but rather a stand-level model. Having accurate predictions of mortality at different levels of planting densities over the entire rotation of a stand is crucial to forest management. They allow us to better determine the timing and intensity of management practices, to facilitate the most optimal outcome for the entity managing the land (Clutter et al. 1983).

The objective for this model was to accurately predict mortality at any time within the stand with a reasonable standard error, less than fifty trees per acre to ensure that the structure of the stand is not changed. This model is designed to be easily integrated with the Stand Management Cooperatives Plantation Yield Calculator. The creation of a model that is easily integrable to others will allow cooperative members to have more accurate growth and yield projections and thus have better information to base their management decisions from.

Chapter 2: Literature Review

There is little doubt that there is a need for quality mortality models to facilitate all kinds of management practices in all forests. It is a well-accepted fact that mortality models are key to accurate long-term growth and yield model projections (Vanclay, 1994). The approaches to modeling mortality however range in approaches and complexity depending on the objectives of the researchers. There is the differentiation of regular versus irregular mortality that was described by Lee in 1971. Regular mortality is classified density dependent mortality;

therefore, it is competition induced suppression that is driving mortality rates in a stand. Irregular mortality refers to larger scale disturbance-based mortality, looking at fire disturbances, pest or disease outbreaks. The type of mortality one is trying to capture will affect the type of model built. In this study, we are modeling regular morality. It is important to note that stand density related mortality, or regular mortality does have the ability to decrease stand-level growth rates and lead to higher susceptibility of disease and pest outbreaks in a stand (Lee, 1971, Vanclay, 1994). So, while regular mortality does not explicitly include disease and pest outbreaks it implicitly does depending on stand density.

Taking a closer look at what kind of models have been developed previously when it comes to stand-level regular mortality models, many of them are based on the power law of self-thinning proposed by Yoda in 1963. This theory is based on the idea that independent of age and site quality, there is a maximum size-density relationship that exists for any stand when the mean plant weight is regressed by the number of trees. This concept led to a multitude of different statistical approaches. From ordinary least squares, principal component analyses, quantile regressions, and stochastic frontier analysis, statistical approaches have been tried through and through to capture this relationship of morality at the stand-level (Zhang et al. 2017). To truly capture the pattern we are looking for, with a certain level of confidence that we are not over or under predicting morality an extensive database is needed and a certain plot size to ensure the correct kind of mortality is being represented. Many of the different approaches have been done using different types of data. The type of data directly impacts the confidence of the model, the best kind of data for this kind of modeling is permanent plot remeasurements (Weiskittel et al. 2011). It allows the modeler to have an extensive database with a wide variety of data to ensure you are capturing the right kind of mortality. The different kinds of data collection that were used were from yield tables (Lee 1971), temporary plots (Hann and Hanus 2001), permanent forest inventory plots (Monserud et al. 1999), research plots (Hann et al. 2003) and remeasurements from forest inventories (Wykoff et al. 1982).

It is also important that the plot size that one is using is capturing the correct kind of mortality that you are looking to model. Plot sizes that are too small can potentially capture irregular morality if an event happened in that area, and thus your entire model would continue to over predict mortality throughout due to this design flaw (Curtis et al. 2005, Weiskittel et al. 2011). Curtis and Marshall recommend that the number of plots in an area are determined to ensure that accurate estimates of stand conditions, mortality and stand increment are met (Curtis et al. 2005).

Chapter 3. Methods:

Chapter 3.1 Study Area:

Data for this analysis originates from the University of Washington's Stand Management Cooperative: a series of spacing trials initiated in the 1980s and 1990s with installations spread throughout Oregon and Washington. The geographic distribution covers eight distinct biogeoclimatic zones as delineated by the United States Geologic Service (Fennemen et al. 1946) (Figure 3.1); the 34 research installations cover a wide range of elevations and site indices across regions. The locations of the installations are showing in Figure 3.1 with critical metrics for each site summarized in Table 3.1. The sites are a wide range of ownership from private to public lands. The sites were planted as pure Douglas-fir (*Pseudotsuga menziesii [Mirb.] Franco)* stands. These are referred to as Type III installations- also known as the spacing trial experiments. While there are installations that do contain other species, the objective was to only focus on the Douglas-fir planted sites.

Region	Count	Region Name	Min Elevation	Max Elevation	Site Index (SI30)
1	1	BC- Victoria	750	750	65
2	1	BC- Mainland	1700	1700	84
3	4	Olympics	330	400	65-96
4	3	WA Coast	400	400	80-93
5	6	Puget	550	1700	43-94
6	1	WA Cascades	750	1700	71-89
7	3	OR Coast	50	1100	67-95
8	4	OR Cascades	100	2000	66-99
9	0	Klamath	NA	NA	NA

Table 3.1 Regional site data for the installations used from the Stand Management Cooperative. Including the number of installations included in each region and their maximum and minimum elevations and average site indexes.

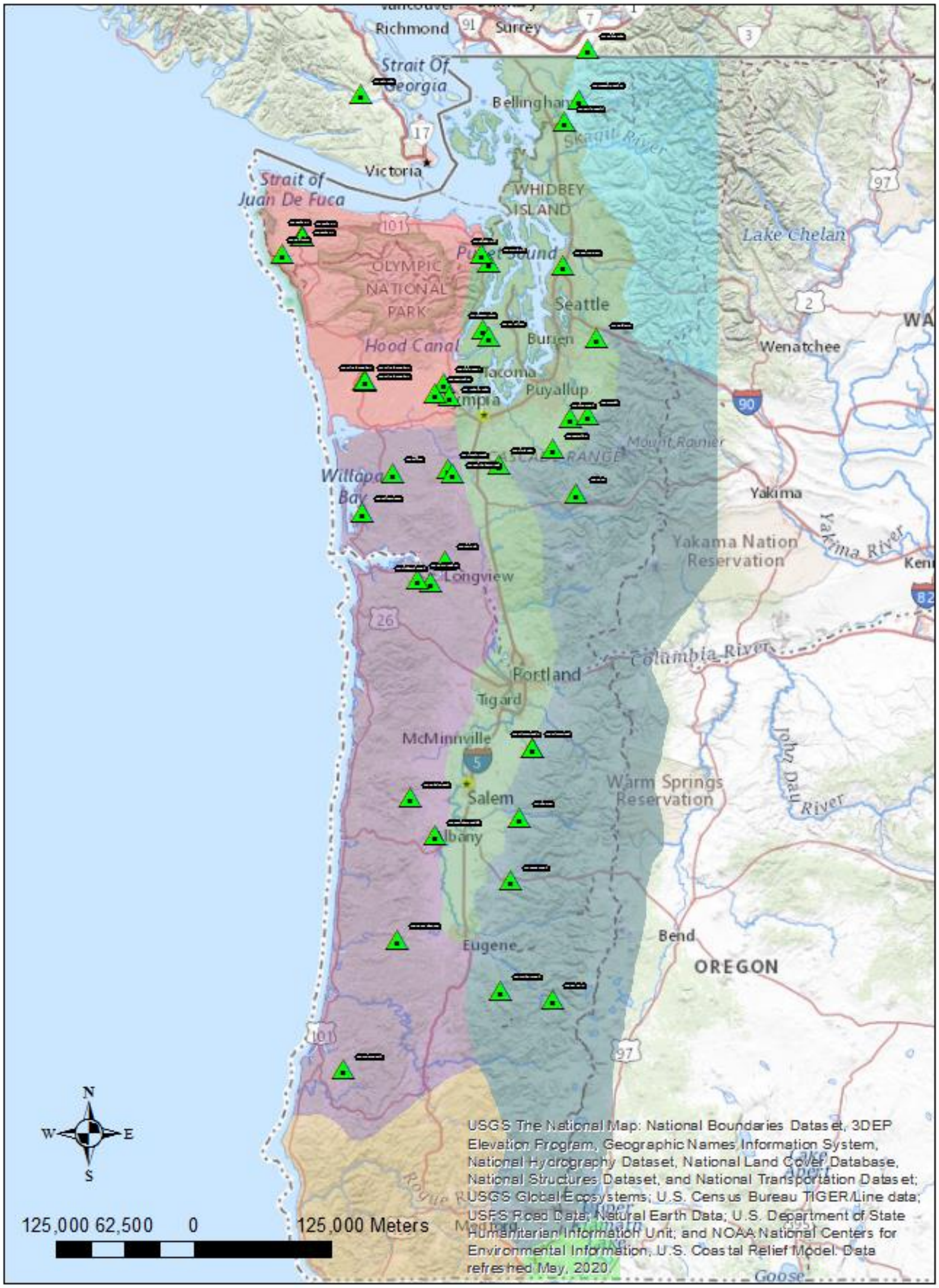

Figure 3.1: Regional map of the type three installations installed by the Stand Management Cooperative, study sites range from southern British Columbia to southern Oregon, all west of the Cascade Mountains.

Chapter 3.2 Study Design and Data Acquisition

Each installation was planted with Douglas-fir (Pseudotsuga menziesii [Mirb.] Franco) as six distinct spacings (densities): 20’x20’ (approx. 100 TPA), 15’x15’ (approx. 200 TPA), 12’x12’ (approx. 300 TPA), 10’x10’ (approx. 440 TPA), 8’x8’ (approx. 680 TPA) and 6’x6’ (approx. 1210 TPA). The diameters of every tree on every plot have been re-measured at four-year intervals. Heights and height to crown-base are measured on a subsample of 42 trees per plot spread across the diameter distribution. A model (Hernandez-Cuevas et al. 2018) relating height to diameter is then fit to predict heights on all remaining trees. All remeasurement data is kept in a database managed by the Cooperative.

In addition to direct measurement data, the SMC database includes metrics calculated from the tree-data: Curtis’s relative density; basal area (per tree and per acre); and height to diameter ratios. This data was augmented with planting records and survival surveys to provide us with accurate information of what each plot in each installation was planted at. Using both six-month stocking records and 2-year survival surveys collected by cooperative members’ own foresters, we were able to assert values backwards to account for the discrepancies of where stocking/survival surveys were not completed to get an informed estimate of what each stand was planted at.

Planting Density	Count	Trees Per Acre	Curtis’s Relative Density	Quadratic Mean Diameter	Site Index
100	217	138	12.6	7.1	84.4
200	214	202	18.5	6.3	80.9
300	215	279	22.7	5.8	79.3
440	193	352	25.8	5.6	79.8
680	191	493	29.2	5.0	77.2
1210	186	745	35.8	4.7	79.6

Table 3.1.2 The count or number of plots included at each planting density, the mean Trees per acre corresponding to each planting density over the entire measurement period, as well as the mean Curtis’s relative density, quadratic mean diameter and site index 30 that all correspond to each starting density and are calculated from the entire measurement period, unless the measurement is static (site index-30).

Chapter 3.3 Response Variable Selection

Two different response variables were considered to capture mortality. The first being a relative response variable called Percent of Planting Density Perished (PPDP). This variable was created to be a percent of how many trees had died at any given time in reference to their initial planting density. It was calculated by taking the initial planting density of the stand subtracting the current trees per acre recorded, dividing it by the original planting density and then multiplying it by 100 to get the final value as a percent. A percent was chosen instead of a proportion, as using proportions can make calculations difficult further down and increases the chance of error in interpretation.

The second variable used was an absolute response variable. It was the cumulative dead trees per acre over time. It was calculated from subtracting the current trees per acre from the original planting density to allow us to have an absolute number of dead trees per acre (dTPA) that have occurred at each time point.

Chapter 3.4 Model Selection

While logistic regression has historically been the commonly used approach in creating a mortality model (Hamilton 1974), there is some work done using exponential functions to capture and predict mortality accurately (Zhang et al. 2017). We used a parameter prediction approach using sigmoidal non-linear curves. Schumacher's non-linear equation (equation 1.1) was the first equation that we used to try to model mortality over time, where time is the regressor. Schumacher was chosen because it was the simplest non-linear, sigmoidal curve with only two parameters to work with.

$$\boldsymbol{y} = \boldsymbol{asymptote} * (\boldsymbol{e}^{-\boldsymbol{rate} * \boldsymbol{x})}) \ (1.1)$$

Where y= response variable

Asymptote= parameter that determines how high the curve goes and where it tops off

Rate= parameter that determines how steep or gradual the slope will be

X= $^{1}/_{age}$

Here y is our response variable, which is either PPDP or dTPA. The asymptote is a parameter that controls how high the Schumacher values are, where they top-out. The rate parameter determines the look of the curve, how quickly or slowly does mortality occur in stands, it determines the slope of the curve (Sit et al. 1994). X in this case is the inverse of age (Mackinney et al. 1937).

The sigmoid curve that this non-linear function takes was ideal to work with as it also coincided with other work done by the SMC and would be easy to integrate into the Plantation Yield Calculator later. Unlike other models we chose to use a non-linear approach to capture the non-linear behavior of mortality in the stand at different starting conditions, as well as create a cohesive model to join the SMC’s plantation yield calculator. Chapman-Richards was also used in the modelling process, when it became apparent that Schumacher’s was not going to be adequate. Chapman-Richards (equation 1.2) is a more flexible four parameter models, that also have the upward sigmoidal curve to them.

$$\boldsymbol{y = asymptote\left(1 - e^{-rate*(x-offset)}\right)^{shape}} \quad (1.2.)$$

Where y= response variable

Asymptote= parameter that determines how high the curve goes and where it tops off

Rate= parameter that determines how steep or gradual the slope will be

Shape= parameter that determines where the inflection point of the curve will be

Offset= parameter that determines when mortality will begin

X= age

Our y variable is the response variable, which is either PPDP or dTPA, our mortality estimates. The asymptote parameter is determining how high the curve goes, and where it flattens out, there is a natural maximum of mortality that the model can predict which is the maximum number of trees planted in a stand. Our rate and shape parameters are determined by a combination of dynamic and static variables. The rate is how steep or gradual the slope is, the shape is when the inflection point of the sigmoidal shaped curve will be (Sit et. al 1994). Our offset parameter is a shifting parameter of when mortality will occur. Will mortality be observed directly after planting or will be occur a set number of years down the line. It is expected that the starting stand densities would have a key role in determining the offset. X in the Chapman-Richards function is age in years.

There has been some work previously done with sigmoidal curves to predict tree mortality, so we felt validated moving forward with it (Somers et al. 1980, Coble et. al 2006). Chapman-Richard's function has been used to relate to mortality man times over the years, as a growth function it can be used to capture the relationships as it does not rely on an intrinsic link to an underlying population process (Gove et. al. 2019). Chapman-Richards four parameters control for different aspects of the model, the *a* parameter is considered to the be the asymptote, which in this case in confined by an upper limit of 100% due to the set-up of the response variable. The *b* and *c* parameters are the rate and shape parameters respectively and the *d* parameter is considered the phase shift.

A principal components analysis was then performed on the right-hand side variables to identify which variables explained most of the variability seen in the data itself (Borcard et al. 2018).

As modeling progressed, it was realized that we had included predictors on the right-hand-side that explained the response variable that would bias the model. The modeling approach was re-evaluated, and it was determined that there were two possible pathways to take moving forward. Either continuing to use with the previously explained response variable of PPDP or changing the response variable to be an absolute variable.

PARAMETER PREDICTORS	DESCRIPTION	TYPE	UNITS
ELEV	Elevation	Continuous	Feet
SI30	Site Index at age 30	Continuous	Feet
SI3012	Square Root of Site Index at age 30	Continuous	Sq. root (Feet)
SDI	Stand Density Index	Continuous	Trees per unit area
PLDEN	Planting Density	Continuous	TPA
PLDEN2	Planting Density Squared	Continuous	Planting Density TPA^2
BA	Basal Area	Continuous	Square Feet
BA2	Basal Area Squared	Continuous	Feet^fourth
R1	Region 1: BC- Victoria	Discrete	Binary
R2	Region 2: BC- Mainland	Discrete	Binary
R3	Region 3: Olympics	Discrete	Binary
R4	Region 4: WA Coast	Discrete	Binary
R5	Region 5: Puget	Discrete	Binary
R6	Region 6: WA Cascades	Discrete	Binary
R7	Region 7: OR Coast	Discrete	Binary
TPA	Trees Per Acre	Continuous	TPA
TPA2	Trees per Acre squared	Continuous	TPA^2
Y2BH	Years to Breast Height	Continuous	Years
CRD	Curtis' Relative Density	Continuous	Continuous

RS	Relative Spacing	Continuous	Feet
QMD	Quadratic Mean Diameter	Continuous	Inches
H40	Height 40	Continuous	Feet
AA, AB, AC	Scaling predictors	Continuous	
LATITUDE	Latitude	Discrete	Decimal degrees

Table 3.4.1 Parameter predictors used in the modelling process. From dynamic to static variables that are thought to be able to describe mortality over time.

After further discussion and exploration, the decision to go forward with an absolute response variable was decided upon, and that variable being, dead trees per acre (dTPA – cumulative over time) was deemed the most appropriate. Scaling predictors were also needed to differentiate the six planting treatments. These new predictors while biologically are not useful, are useful in the sense that they can differentiate these treatments without TPA being needed. Originally in the model we had both TPA and PLDEN informing the right-hand side as well, which was incorrect since the response variable is a function of those two predictors, so in order to adjust for this TPA needed to be removed from the set of predictors, this then removes many other predictors originally included since they were dependent/ derived from TPA, these variables included, CRD, RS, and QMD. We were left working with H40, BA, Regions, and Elevation. So, to create the scaling vectors we looked at combinations of H40, BA and PLDEN to see how they could be related to create the differentiation needed. It was determined that there were three potential combinations that showed the relationship we were looking for well and decided to move forward and add those to the model to determine how it would improve the model fit.

What variables used were determined by the type of model that was created. The first model attempts were executed using a relative model but once the determination of using an absolute model was discussed and explored the realization that some of the variable, we were trying to use would not always be available to us to inform the model later down the line and the types of variables that are currently available to work with (Table 3.4.1.)

Chapter 3.5 Computed and Site-Specific Variables

Site Specific Variables

Elevation is an absolute predictor as it is site specific and means it will not change over time, it is an absolute predictor. Elevation coupled with Latitude should give us a good idea of how the physical location of each site is affecting mortality at the site level. Between these two and the regional data that we have included in form the USGS, we should have adequate site-specific data that can inform the model to capture the site-specific changes. These all act as a proxy for soil quality, parent material and climatic information that is not specifically accounted for.

Trees Per Acre

Trees per acre is also an absolute measurement in the sense that it is not dependent on other variables to be gathered to deduce the number, it is truly the number of individual trees in a unit area. It is useful when regarding regeneration purposes and examining stands for specific tree characteristics.

Quadratic Mean Diameter

Quadratic mean diameter the diameter corresponding to the tree of arithmetic mean basal area (Curtis et al. 2000). It can be calculated by the following equation (1.4)

$$\overline{D_q} = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n}} \text{ (1.4)}$$

Where d= diameter at breast height

n= number of trees measured

Basal Area

Basal area is an absolute measurement, which gives the total cross-sectional area of stems at breast height. Basal area is often used for volume growth but is does not account for the number of trees on the unit area (Tappeiner et al. 2015). Basal area is used as a measure of density. This is calculated by using the equation shown below (1.5).

$$.005454154 \sum_{n} d_i^2 \text{ (1.5)}$$

Where d = diameter at breast height

Top Height

Top height or height forty is the average height of the forty largest trees per acre, based on diameter at breast height. This measurement is only useful once a stand has been established enough that the trees are past breast height.

Relative Spacing

Relative spacing is the average spacing of trees related to the average height of trees. At maximum stand densities, Douglas-fir exhibit a nonlinear relationship between individual tree height and diameter, it can also be regarded as a function of the ratio between the measured number of trees per unit area and the maximum possible with average height (Tappeiner et al. 2015). Relative spacing indices are often based on an observed number of trees per unit area relative to stand average height, therefore can be calculated from the either of following equations depending on whether you have average height or average DBH (equation 1.6 & 1.7)

$$RS = \frac{\sqrt{A/_N}}{H_D} \quad (1.6)$$

Where A = unit area

N = number of trees per unit area

H_D = average height of dominant trees

Or

$$RS = \frac{\sqrt{A/_N}}{D_D} \quad (1.7)$$

Where A = unit area

N = number of trees per unit area

D_D = average dbh of dominant trees

Curtis's Relative Density

Curtis's relative density is a measure of competition on the stand and can be calculated by any combination of two basal area per unit, quadratic mean diameter and the number of trees per unit area (Tappeiner et al. 2015). It is comparable to Drew and Flewelling's relative density index, which is used to relate trees and stand volume, while also considering height, it depicts stand development and volume over time at different densities. However, CRD can be misleading as the value calculated informs the user only so much, but the value is not indicative of stand structure, CRD is a useful competition index but a 680 TPA stand and a 440 TPA stand the way that the competition looks could differ greatly (Curtis, 2010). The stand dynamics, the way and rate at which the stand moves through growth periods, are very different over time, but the CRD value could be the same. CRD can be calculated by the following equation (1.8)

$$CRD = \frac{BA}{\sqrt{D_q}} \quad (1.8)$$

Where BA = basal area

D_q= quadratic mean diameter

Scaling Predictors

We have three scaling predictors that we created, AA, AB and AC that are used to differentiate among the starting spacing conditions described in the previous sections. These scaling predictors are combinations of readily available predictors within the model that allow there to be obvious differentiate among the starting planting densities. These variables are combinations of top height, basal area and planting density (equations 1.9, 1.91 & 1.92).

Combinations of other variables were considered by plotting the combinations of them against total age, which is the regressor, but these three differentiated the starting conditions of the stands the best (figure 3.5.1).

AA = H40* BA (1.9)

AB = H40* PLDEN (1.91)

AC = BA * PLDEN (1.92)

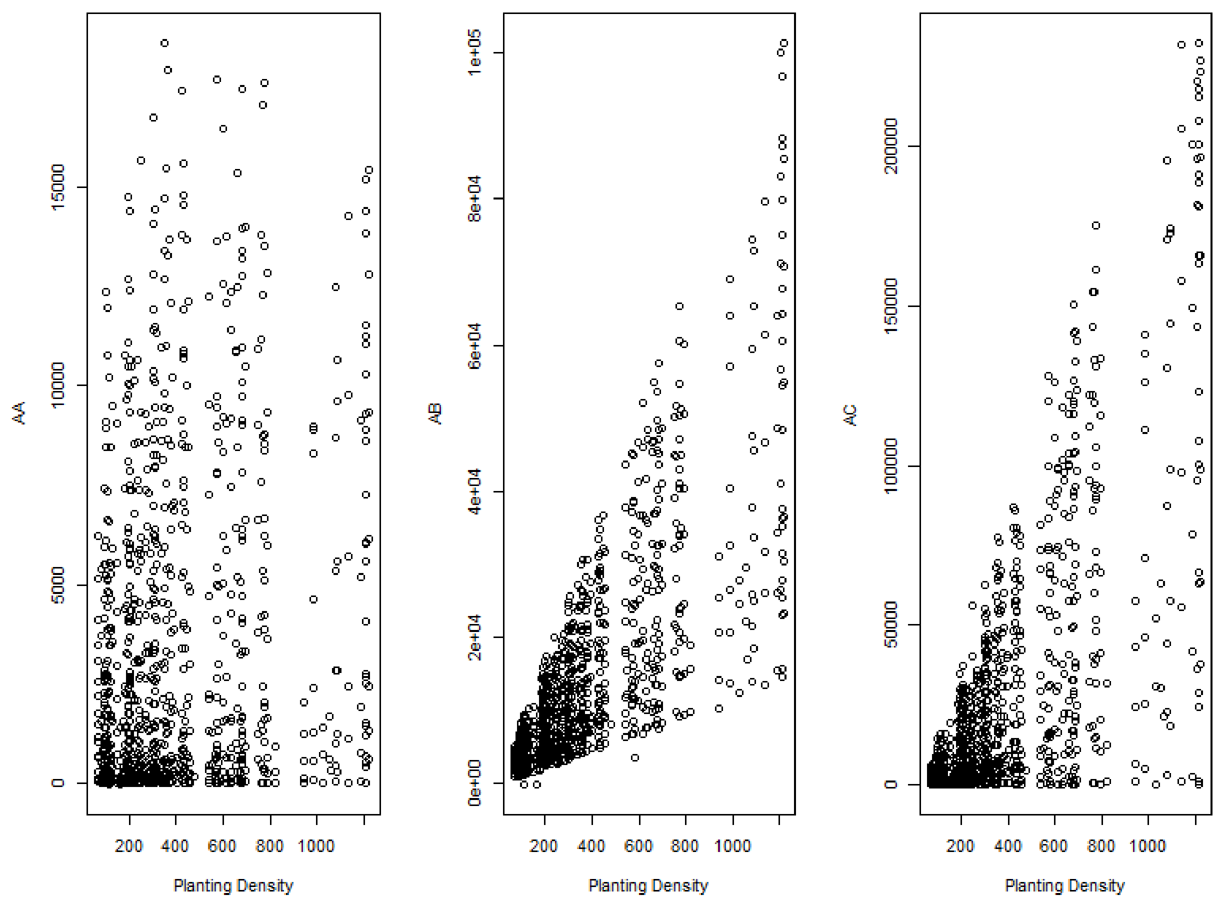

Figure 3.5.1 Scaling predictor created to allow the model to differentiate among planting densities, compromised of combinations of height-forty, planting density and basal area. AA = H40 BA, AB = H40* PLDEN, AC = BA * PLDEN. Used in the modeling process to help the model differentiate predictor values by associating them with a specific starting planting density.*

Chapter 3.6 Model Validation:

A bootstrapping process was performed for validation of the parameter prediction approach. Bootstrapping is an arbitrary statistic technique used to determine if the statistics are sensitive to a few outlying values. It is used to estimate the bias of an estimator and works by repeatedly selecting random observations from the data and recalculating the statistics (Adler 2009). Since we were using a reduction approach the model was re-run 100 times through the bootstrapping procedure to ensure that the parameter chosen as least significant was indeed the one that was impacting the model the least. This process was run for every model output until we deemed the model to be as accurate and reduced as possible. This was determined by evaluating R^2 values and p-values. The parameter predictor with the highest p-value (<.2) was eliminated from the process as it was not adding to the model's predictive power but adding noise to the model instead. When we deemed that the model was finalized, we visually assessed the residuals to allow us to make any assumptions of any patterns if we were seeing them.

Chapter 4. Results

Plotting each possible response variables against our regressor variable (age), for each starting planting density gave confirmation that there is a slight upward trend to most of the data and that a non-linear function would be the correct approach (figure 4.1). This figure shows response variables (PPDP, dTPA) segmented by planting density.

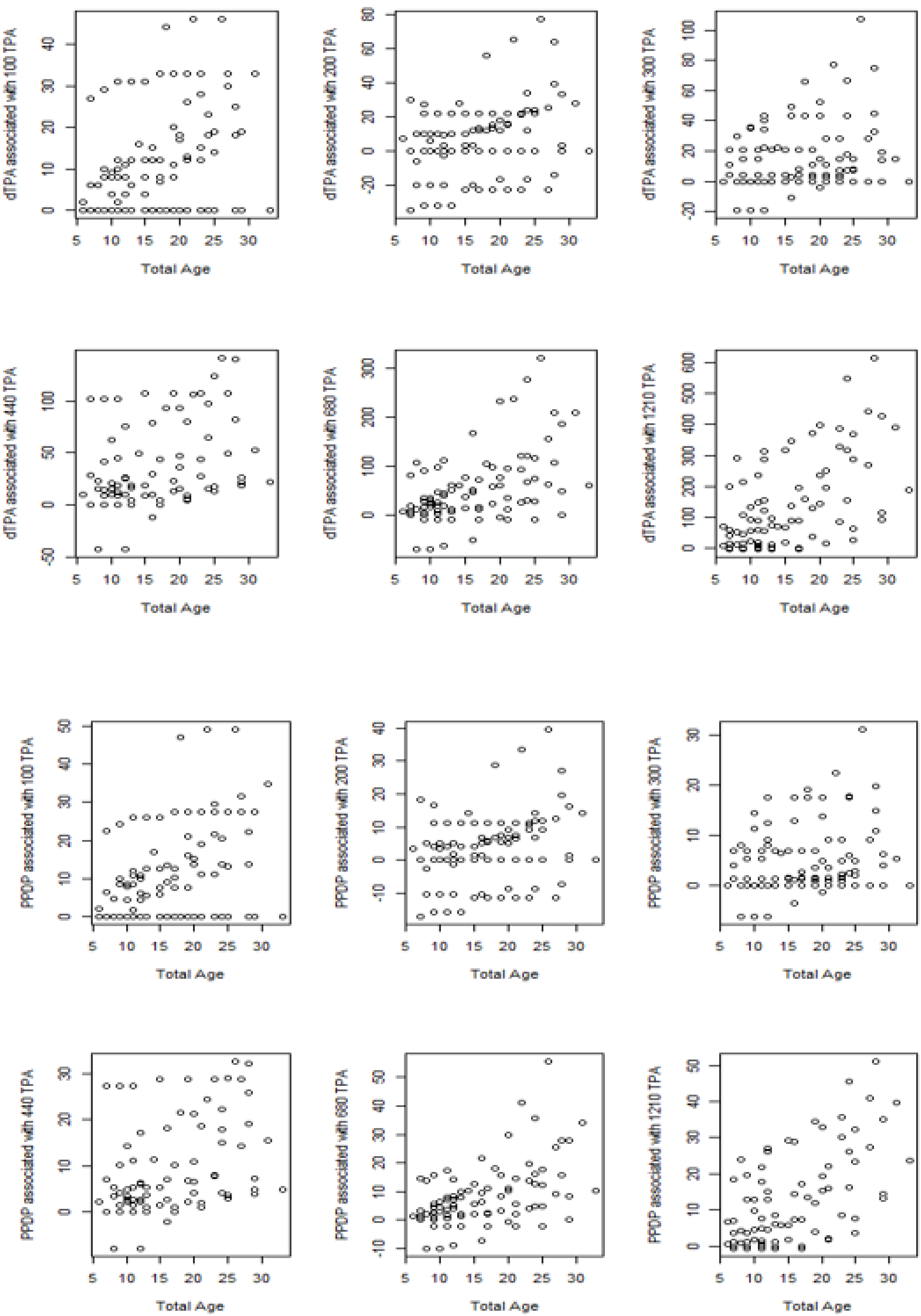

Figure 4.1. Response variables dTPA and PPDP at each starting planting density plotted over time, used to observe the shape of the behavior in order to ensure that the desired sigmoidal shape is present in the data.

Chapter 4.1 Schumacher's

Schumacher's non-linear equation was the first attempt at modeling mortality in the type three installations and used as a preliminary analysis of the data. Schumacher's equation form includes only an asymptote and a rate parameter, not a shape parameter. Even when informed by several predictors, the two-parameter model demonstrated itself inadequate to predicting mortality. The parameters were estimated as linear combinations of several predictor variables which allowed visually to tell that the predictive model was not capturing the trend of mortality as expected. Figure 4.1.1 demonstrates that the model is underpredicting mortality, capping PPDP at 15%, where it is known from the original dataset that mortality rates were indeed high than that. This allowed us to consider different non-linear models that also have the sigmoidal behavior to them, that may have more flexibility in them by having more parameters to capture mortality.

Given the lack of performance for PPDP, it was not considered/included any further in the analysis.

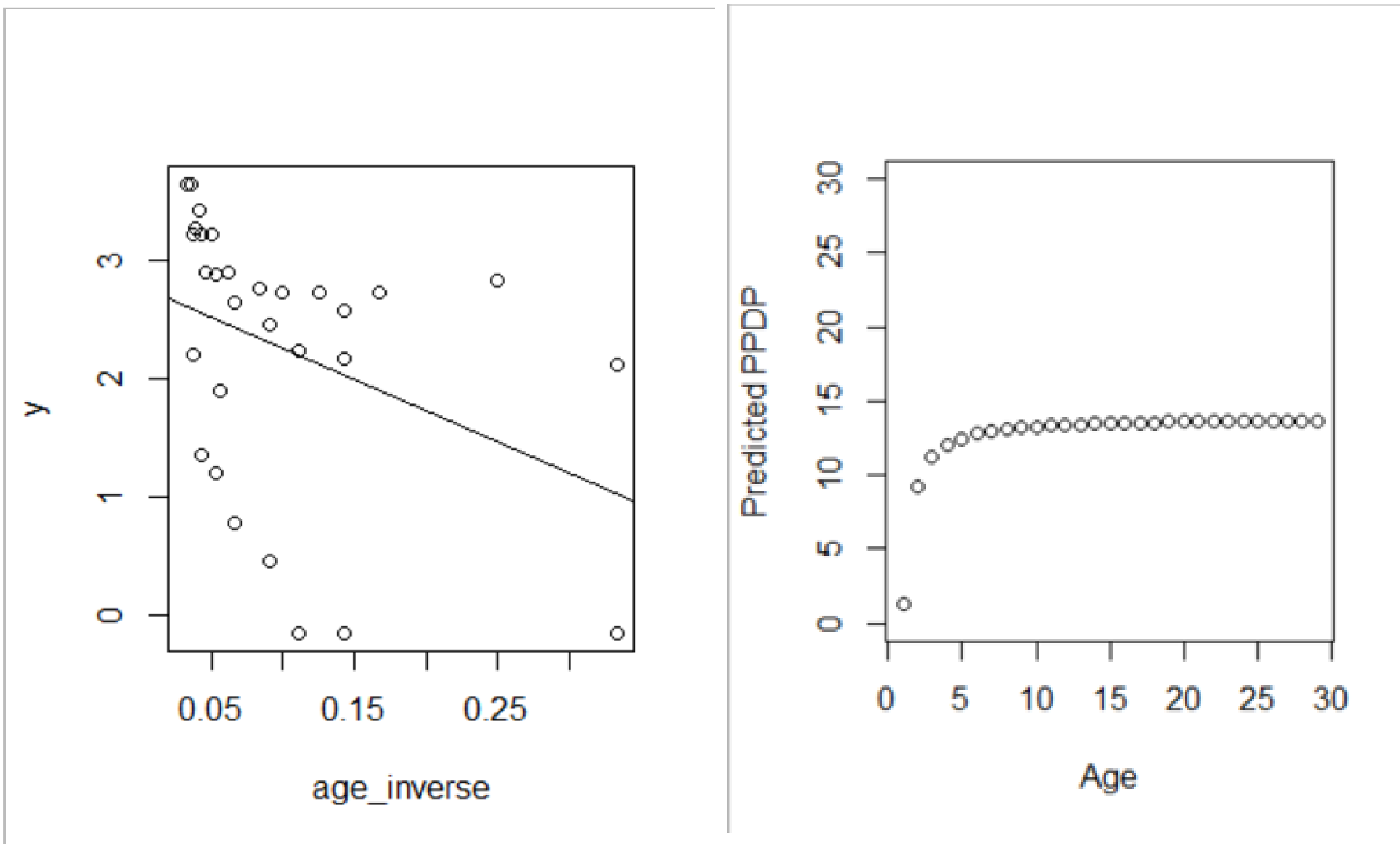

Figure 4.1.1 Linearized Schumacher's from model fits after the reduction process was completed and the predicted PPDP from the model reduction process, you can clearly see the data bifurcate, indicating Schumacher's is an inappropriate fit. The figure on the right shows that PPDP has an asymptote at 15%, which given the actual known PPDP values is underpredicting mortality.

Chapter 4.2 Eigen Analysis

An eigen analysis was done prior to starting the reduction process of model using the Chapman-Richards non-linear function. This allowed us to fit the model with key variables that were identified in the eigen analysis (table 4.2.1) These variables were then put into the shape and rate parameters only, as we were able to identify a set of predictors for the asymptote and offset parameters as they were more objective than the rate and shape parameters. Since the asymptote and offset parameters are the parameters that control for the asymptote and phase shift respectively, they have a more distinct set of parameters. The rate and shape parameters are the key ones to use the eigen analysis outputs in to feed the model.

Parameter Predictors	Variable importance					Total	Rank
PLDEN	0.0000	0.0121	0.0000	0.0000	0.0000	0.0121	8
ELEV	0.1828	0.0000	0.0638	0.0228	0.0000	0.2694	2
TPA	0.0000	0.0200	0.0000	0.0530	0.0244	0.0973	5
SI30	0.0000	0.2650	0.0000	0.0000	0.0000	0.2650	3
Y2BH	0.0000	0.0000	0.0495	0.0000	0.0000	0.0495	7
H40	0.0000	0.0000	0.0000	0.0000	0.0048	0.0048	9
CRD	0.1902	0.0000	0.0000	0.0000	0.0000	0.1902	4
RS	0.0000	0.0000	0.0000	0.0578	0.0259	0.0837	6
BAA	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	10
QMD	0.2712	0.0000	0.0981	0.0000	0.0000	0.3693	1
LCR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	11
HD	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	12

Table 4.2.1 Eigen Analysis variable importance output for the original set of predictors used in modeling PPDP with Schumacher's equation. The top five most important variables to explaining the variation seen in the dataset were used in the modeling process with Schumacher's equation. Refer to table 3.4.1 for each parameter predictor definition.

Chapter 4.3 Relative Model

4.3.1 Chapman-Richards with PPDP

We originally used Chapman-Richards with the rate and shape parameter filled with the variables identified in the eigen analysis. A realization that the response variables were also in the right-hand side which then was biasing the model output, led to a change in the approach and what variables were available to use to in the right-hand side. In this process many of the right-hand side predictors had to be removed as they were derived from or dependent on the TPA measurement, which had to be removed as it is also in the PPDP response variable. This left the model with limited predictors informing the response.

4.3.2 Chapman-Richards with PPDP with additional predictors

To correct this, new predictors were created that were designed to be added to the right-hand side to scale and differentiate the different starting planting densities, as the predictors left were not capturing the behavior expected. You can see in figure 4.3.2.1, that there is major over and under prediction occurring in this model. While the reduction process has the possibility to clean up some of the noise observed here, it was not the most promising initial fit (figure 4.3.2.1). The model statistics after running were not promising as well, the R^2 value was .4603, meaning that the model was only explaining 46% of the variation in the data. The mean annual error (MAE) was 5.603 and the standard deviation (SD) was 7.94. So, while the SD and MAE values were fine alone, considering the actual model performance, it was not adequate to move forward with.

Final set of predictors used to estimate parameters: PLDEN, PLDEN2, BA, BA2, SI30, SI3012, REGIONS (1-7), AA, AB, AC, ELEVATION

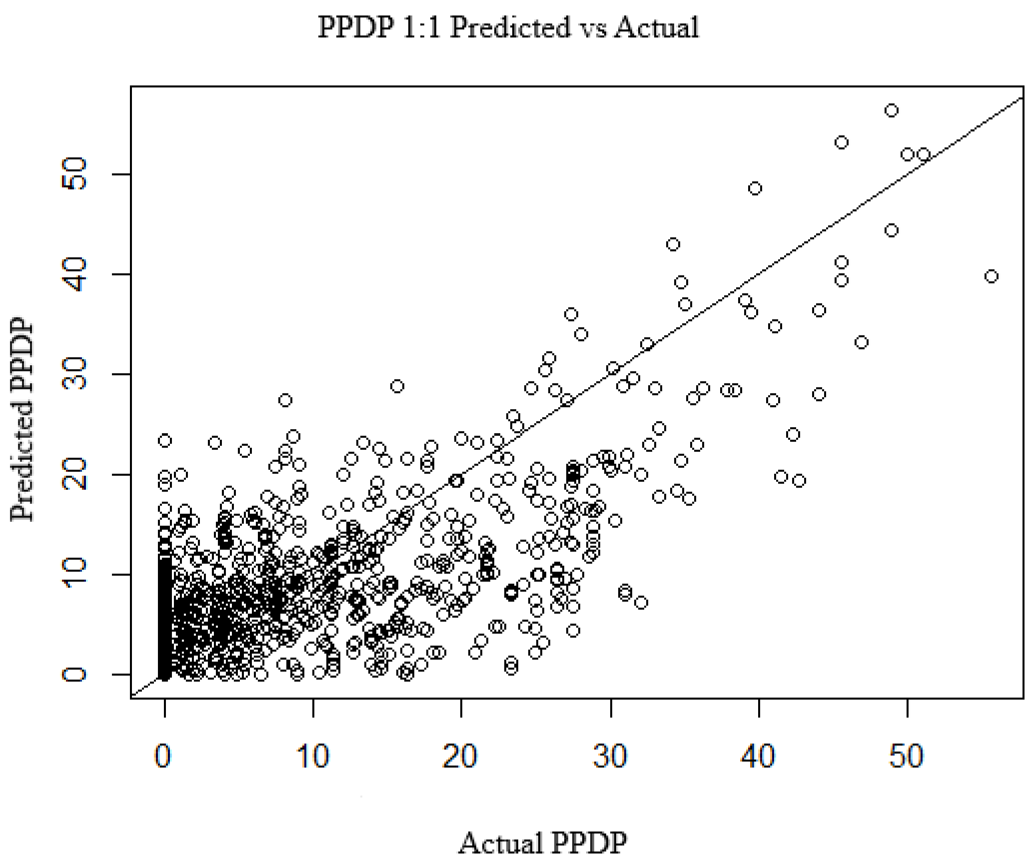

Figure 4.3.2.1 One to one model fit with PPDP as the response and the new set of predictors informing the parameters. This shows the over- and underprediction of the model at each planting density.

Chapter 4.4 Absolute Model

4.4.1 Chapman-Richards with dTPA

After reworking the model with an absolute response variable, initial fits of the two types of models, absolute vs. relative were compared and it was determined that an absolute model fit the data much more precisely. This was determined by comparing R^2 values and visually assessing the one-to-one fits. The data cloud was much more concise and appeared to be a better starting point than when we were working with a relative model (figure 4.4.1.1 vs figure 4.4.1.2). The reduction process was then performed for two versions of the absolute model and the fit was examined every few runs to ensure that the model was fitting correctly. The two versions were a bounded asymptote and an unbounded asymptote. The unbounded asymptote appears to be the better fit as the reduction process continues. The two models were compared using their coefficient of determinations (R^2), mean annual errors (MAE) and standard deviations (SD), these values and each parameters inputs are summarized in table 4.4.1.

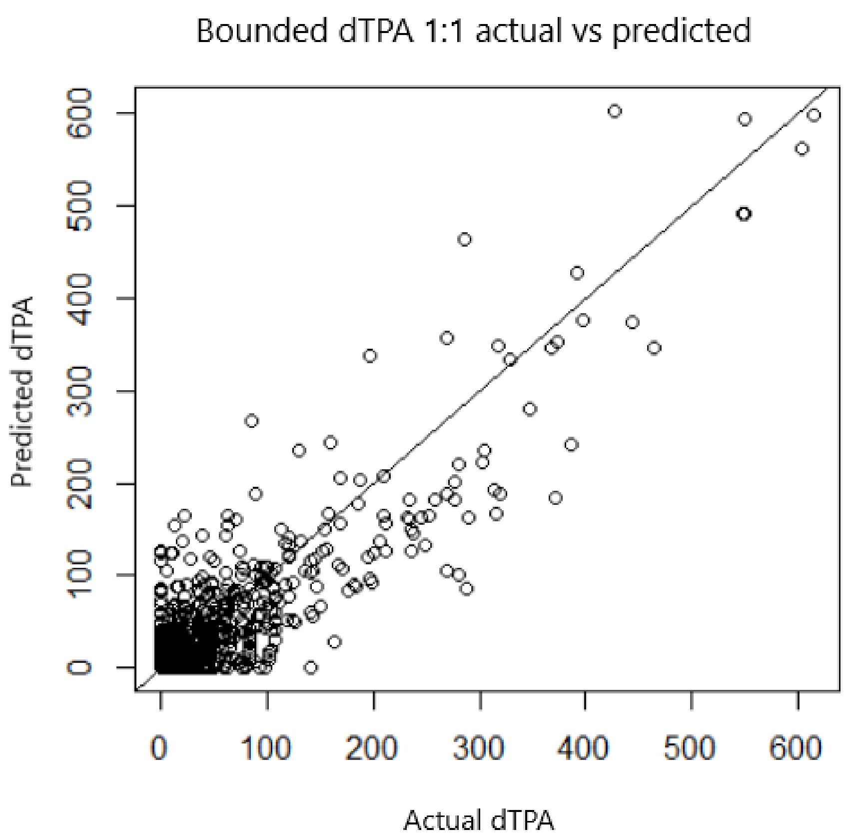

Figure 4.4.1.1 One to one model fit with dTPA as the response variable- bounded asymptote initial model fit. The asymptote parameter is bounded at 750, meaning that no more than 750 dTPA can be predicted.

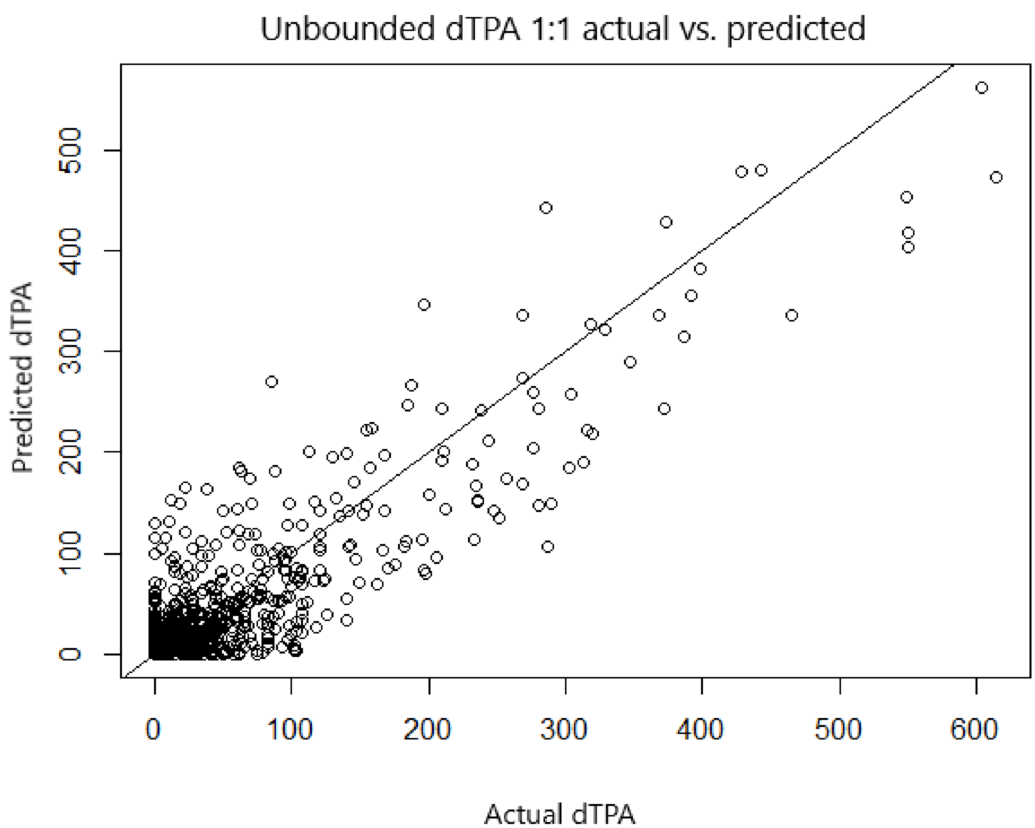

Figure 4.4.1.2 One to one model fit with dTPA as the response variable- unbounded asymptote initial model fit. There is still some over- and under prediction being seen in the initial model fit of the data.

	Bounded Model	Unbounded Model
R^2	.79	.78
SD	35.17	36.87
MAE	21.77	22.20

Asymptote parameter predictors	PLDEN, PLDEN2, R4, R6, R7	PLDEN, PLDEN2, R1, R2, R4, R5, R6
Rate parameter predictors	PLDEN, R1, R2, R4, R5, R6, R7, ELEVATION, AA, AB, AC, SI30	PLDEN, R1, R2, R4, R5, R6, R7, ELEVATION, AA, AB, AC, SI30, SI3012
Shape parameter predictors	PLDEN, R1, R2, R4, R5, R6, R7, ELEVATION, AA, AB, AC, SI30, SI3012	PLDEN, R1, R2, R4, R5, R6, R7, ELEVATION, AA, AB, AC, SI30, SI3012
Offset parameter predictors	PLDEN, PLDEN2, R1, R2, R4, R5, R6, R7	PLDEN, PLDEN2, R1, R2, R4, R5, R6, R7

Table 4.4.1 Final Statistics of the model reduction process in both the unbounded and bounded asymptote models. The bounded model means the asymptote cannot surpass 750 dTPA. Including information on the R^2, SD and MAE values, as well as the predictors left in each parameter at the final output.

Both models performed relatively similarly statistically, with 78% and 79% of the variation being seen in the data captured by the model. Looking closer at the standard deviation, it is larger than preferred with two-thirds of the data being within either a 70 or 72 trees per acre band. However, looking at the mean annual error value, which is also known as the predictive error of the model, you can see that respectively there is a 42 and 44 trees per acre band around the predictive value. Meaning that for any predicted value there is a chance of being off my less than 50 trees per acre. This is crucial as that level of precision allows us to ensure that stand structure is not being mis portrayed. The final unbounded and bounded models are summarized by each parameter and their coefficient (table 4.4.2, 4.43).

Unbounded model final equation:

$$\boldsymbol{y = asymptote\left(1 - e^{-rate*(x-offset)}\right)^{shape}}$$

Parameter Predictor	Asymptote	Rate	Shape	Offset
Intercept	2.34e02	1.0e-02	1.7e00	4.3e-03
PLDEN	-1.3e-05*	5.7e-03*	1.4e-03*	-6.3e-03
PLDEN^2	1.9e-04*	-	-	-1.1e-05
ELEV	-	-4.8e-06*	-4.6e-04*	-
R1	9.3e01	1.3e-02	2.7e00	-5.8e-03*
R2	-	1.1e-04*	7.2e-01	5.3e00
R4	-3.7e-03*	1.5e-04*	5.9e-01	1.3e01
R5	3.5e02	1.7e-03*	8.9e-01	6.3e-01
R6	1.4e02	2.1e-02	1.9e00	7.1e00
R7	-	2.5e-02	1.2e00	1.1e01
SI30	-	-2.2e-05*	-3.5e-03*	-
SI3012	-	-1.0e-04*	4.7e-03*	-
AA	-	-5.9e-07*	-	-
AB	-	1.4e-06*	-	-
AC	-	-4.0e-07*	-9.0e-06*	-

Table 4.4.2 Final unbounded asymptote model parameter prediction coefficients. If no coefficient is included that parameter was not kept in that parameter in the final reduction process.
** Indicates predictors that could possibly be adding noise to the model-- further reduction would be needed to determine.*

Bounded model final equations:

$$y = asymptote\left(1 - e^{-rate*(x-offset)}\right)^{shape}$$

Parameter Predictor	Asymptote	Rate	Shape	Offset
Intercept	2.0e02	1.0e02	2.8e00	1.9e01
PLDEN	4.9e-02*	-1.9e-06*	4.6e-03*	-9.4e-02
PLDEN^2	-1.7e-04*		-	-8.1e-05*
ELEV	-	-2.9e-07*	-1.8e-04*	
R1	-	1.6e-04*	2.6e-02	7.0e00
R2	-	-9.0e-05*	7.4e-01	4.8e00*
R4	5.1e01	-5.9e-05*	4.5e-01	4.4e01*
R5	-	2.9e-04*	6.9e-02*	5.0e-03*
R6	3.6e01	4.8e-04*	4.2e-02*	4.3e00
R7	3.1e-03*	-1.2e-04*	4.7e-01	6.3e-01*
SI30	-	-4.0e-05*	-3.0e-02*	-
SI3012	-		9.5e-02*	-
AA	-	2.0e-07*	1.1e-04	-
AB	-	1.0e-07*	1.7e-05*	-
AC	-	-2.5e-08*	-8.7e-05	-

Table 4.4.3 Final bounded asymptote model parameter prediction coefficients. Where the asymptote cannot surpass 750 dTPA. If no coefficient is included that parameter was not kept in that parameter in the final reduction process.
** Indicates predictors that could possibly be adding noise to the model-- further reduction would be needed to determine.*

The final model form of the unbounded model was taken and then used to predict dTPA over 35 years and was overlayed onto the average values for each age of the original data for three different starting densities, 200, 440 and 680 TPA (figure 4.4.1.3). For each predictor included in the model (site index, regions, elevation, planting density and scaling predictors), the average values of those predictors were taken at each age for all plots and were used to predict dTPA (see Appendix A.2 Supplementary Material). This was done to show the models predictive power in comparison to the original data.

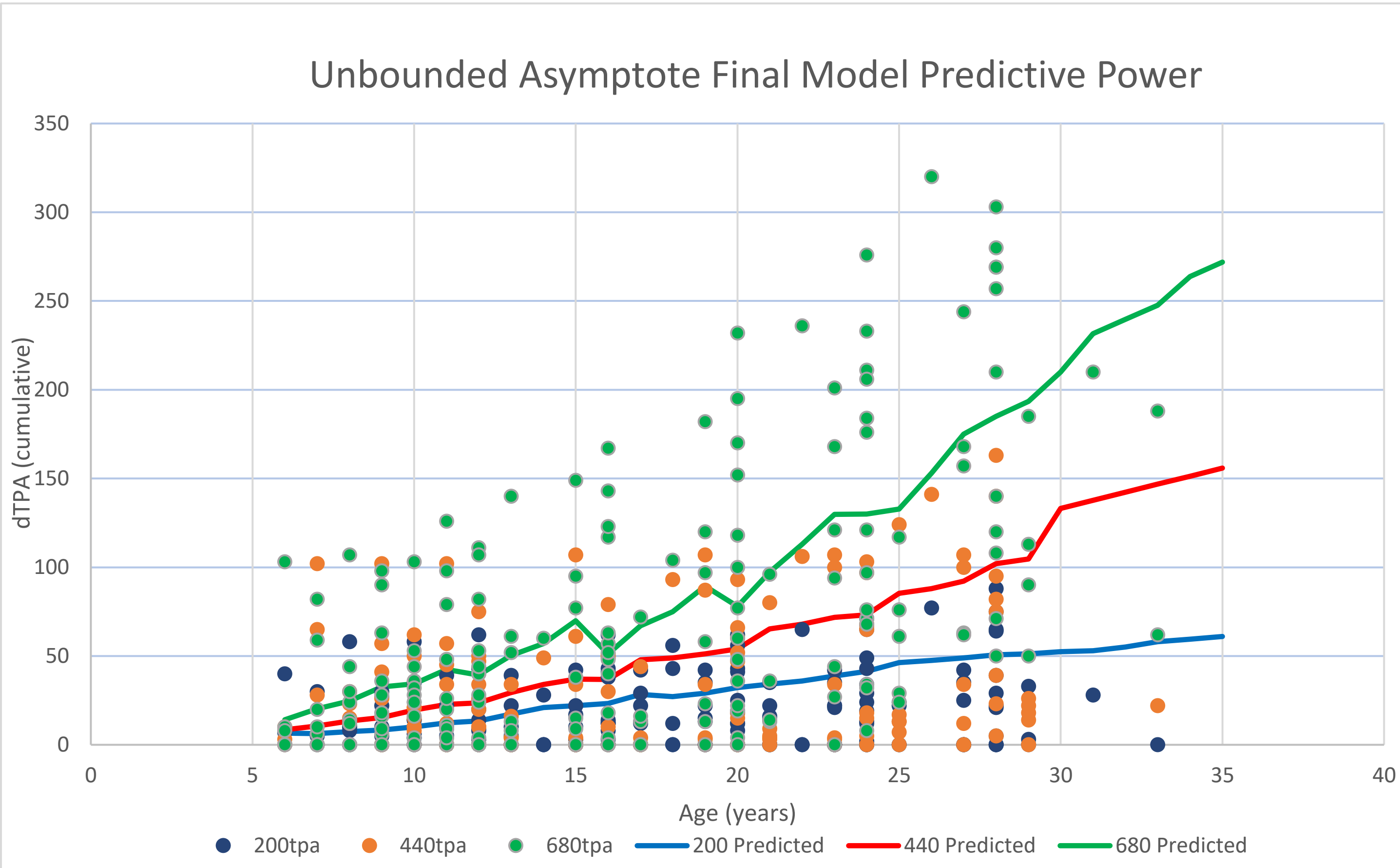

Figure 4.4.1.3 Final Model Predictive Power of the unbounded asymptote model. The final unbounded equation was used to predict the number of dead trees per acre over 35 years. These predictions were overlayed onto the original data.

Chapter 5. Discussion and Model Comparisons

Chapter 5.1 Model Comparison

Relative variables are variables that are dependent on different measurements and are likely to change drastically over time. However, throughout this process it was determined that variables like relative spacing and relative density were not at our disposal given the variables we could work with, they required us to have certain measurements before time that we did not have access to and would not be able to predict further. A relative model was not expressing the relationships we were seeking well enough to move forward that a absolute model would capture. The two types of models came down to the trade-off of accuracy vs. precision. If the decision to continue to use a relative model instead was made, we could potentially be allowing large swings in the quadratic mean diameter predictions which we need to be accurate as we do not have other metrics that allow us to differentiate stand structure in a relative model. Even with the addition of our later added differentiation variables, AA, AB and AC, relative predictors were not capturing the known behavior. In comparing the basic model fits of the relative model, it was visually possible to see the model's shortcomings. It was considered that there was a possibility of missing predictor variables. More site-specific data could have tightened up the initial fit, but after gathering soil and site-specific data for each installation, there were too many inconsistencies in the type of data that was available for our desired sites. Getting the initial fit back of the absolute model and seeing how tight of an initial one to one fit we got, it was an easy decision to permanently make the switch to an absolute model instead. By switching to an absolute model, we also no longer had to worry about the issue of accuracy and precision that could have been lost in the relative model. In the relative model we were using predictors such as relative density to differentiate between our planting densities, however examining that assumption that was made, it was determined that different planting densities can have the same relative density value but indicate very different stand structures.

The decision to move forward with two absolute models, a bounded asymptote at 750 TPA and an unbounded model, was one of curiosity to see which would perform best. An unbounded model will not requiring any re-fitting of the model at any point based on expected mortality, however a bounded model, if a landowner has a region and site index that high mortality is common, and a 750 bound may not allow accurate predictions the model would need to be re-fit in that process. With both models being quite similar in their predictive abilities, it is a decision of whether to take the chance using the bounded model or not. This is not to say that these are the most concise and parsimonious models out there. The reduction process was stopped due to time constraints, and could be further reduced, as the bootstrap final p-values of each predictor indicate that there may be one or more predictors still included in the model that may be contributing noise to the model.

Taking the unbounded asymptote final model final form and using it to predict dTPA over time and overlaying those predictions to our original data was done to see the model's predictive power. The unbounded final model was chosen instead of the bounded model to predict forward due to the potential issue of the model's need to be re-fit if a manager ever expected greater than 750 dTPA during the rotation. So, the unbounded model was chosen to move forward with in this last step to avoid future complications. There are dips and spikes over the predictive power in the model, they correspond to specific data entries where plots within an installation had much lower or higher initial planting densities then the other plots at other installations. The dips and spikes correspond to the model adjusting for these differences in the dataset itself.

Chapter 5.2 Discussion

The purpose of this model is to fill in the gap that the timber industry has right now because most growth and yield models are underestimating mortality. Simulation models like

FVS and ORGANON and many in house ones like the SMC's PYC all underestimate mortality. There are adaptive ways to try to account for that, like in FVS you can perform a manual adjustment to account for this underestimation of morality but that does require a manager to have a tree list to do the adjustment with (Keyser 2008). What we have been able to create is a mortality model that is able to accurately account for morality over a range of starting densities and over its complete lifetime (figure 4.4.1.3). The power to do this without having a tree list makes this much more accessible and useful for forest managers. This removes a barrier that is otherwise in place. By doing this it is allowing forest managers to have more accurate and precise yield estimates and be able to predict more accurate estimates based on the planting regime they deem is appropriate for their site. By only needing a few specific measurements that are easily accessible to foresters, they can improve their estimates.

Implementation of the model is key here. The goal is to create a model that will then inform the Plantation Yield Calculator (PYC) that the SMC has also developed. The PYC is then accessible to anyone within the SMC and can and is used to gather growth and yield data based on several assumptions that the cooperative members can specify. The dataset it assessable to them at all times as well.

Limitations do include that is a Douglas-fir only model. Past that it is specifically for the Pacific Northwest, partial to west side of the cascades as well. It is not globally or even applicable to other parts of the country. However, the process can be replicated for any species in any regions given the same set of parameters. The model is only designed to work in simple, even-aged, planation style forests (VanClay, 1994). To try to apply this directly to un-even aged forests or multiple species, it will not transfer well. In those cases, there are many other relationships that are not being accounted for and a new set of predictors would be required to accurately capture the trend of mortality seen in those systems.

Chapter 6: Conclusions

This work starts the process of increasing accuracy in growth and yield models by incorporating a full Douglas-fir mortality model into growth and yield models by having created a mortality model with non-linear sigmoidal behavior that can easily be incorporated into the SMC's plantation yield calculator. The process was not linear, and much was learned along the way. From trying multiple different non-linear curves, to trying relative and absolute models, to running an eigen analysis, many different techniques were tried, failed, and re-worked. There is much room for future work when it comes to morality models, and this was just the start. Expanding the species that the model works for would be a good starting point, Western hemlock is also a major species in the pacific northwest that could be modeled. However, the behavior and growth differ as western hemlock is much more shade tolerant species, you would expect to see a general different shape to a mortality curve. There is also the possibility for this to be expanded even further, many different models in the past have focused on things like pest or disease outbreaks but is there a general model that could take that risk into account by including a risk factor of spread dependent on stand density, since stand density is often a major factor in the spread and severity of an outbreak in many cases.

Application of this model and the possibility for it to be expanded on is great, as seen with the possibilities listed above. The work will never be perfect and complete as ecosystems are forever changing and with the threat of climate change and zones changing, the processes we are seeing may have to be adjusted with time. We tried to keep the model predictors as simple and the measurements and information needed to feed the model as simple and easily accessible as possible, but by doing that we may have also missed stand metrics that could have informed the mode even better. However, by not exploring these options we are hoping to make it useful for people at every level. The metrics you need are done by easy, quick field measurements, and by easily accessible online data, that has no barriers to entry.

Appendix.1 References

Adler, J. R in a Nutshell. O'Riley Media, Inc. 2009.

Bigler, C., O. U. Bräker, H. Bugmann, M. Dobbertin, and A. Rigling. 2006. Drought as an inciting mortality factor in Scots pine stands of the Valais, Switzerland. *Ecosystems* **9**:330–343.

Brocard, D., Gillet, F., Legendre, P. Numerical Ecology with R. 2018. Chapter 5: Unconstrained Ordination. 153-171.

Clutter, J., Fortson, J., Pienaar, L., Brister, G., Bailey, R. Timber Management: A Quantitative Approach. John Wiley & Sons. 1983.

Coble, D.W., Lee, Y.J., Use of a generalized sigmoid growth function to predict site index for unmanaged loblolly and slash pine plantations in east Texas. 2006. Gen. Tech. Rep. SRS-92. Department of Agriculture, Forest Service, Southern Research Station.

Curtis, Robert O. 2010. Effect of Diameter Limits and Stand Structure on Relative Density Indices: A Case Study. West. J. Appl. For. 25(4).

Curtis, Robert 0., Marshall, David D. 2000. Why Quadratic Mean Diameter? Pacific Northwest Research Station. West. J. Appl. For. 15(3):137-139

Curtis, Robert O., Marshall, David D. 2005. Permanent-Plot Procedures for silvicultural and Yield Research. Pacific Northwest Research Station. General Technical Report PNW-GTR-634.

Fennemen, N.M., Johnson, D.W. Physiographic divisions of the conterminous U.S. 1946. U.S. Geological Survey

Gove, J. Lynch, T., Ducey, M. The Chapman-Richards Distribution and its Relationship to the Generalized Beta. *Forest Ecosystems.* 6:27. 2019.

Hamilton, D. 1974. Event probabilities estimated by regression USDA Forest Service Research Paper. INT 152.

Hann, DW, and ML Hanus. 2001. *Enhanced Mortality Equations for Trees in the Mixed Conifer Zone of Southwest Oregon.* Research Contribution 34, Forest Research Laboratory. Oregon State University.

Hann, DW. Marshall, D.D., Hanus, M.L., 2003. *Equations for predicting height-to-crown base, 5 year diameter growth rate, 5-year growth rate, 5-year mortality rate, and maximum size density trajectory for Douglas-fir and western hemlock in the coastal region of the Pacific Northwest.* Research Contribution 40. Oregon State University. College of Forestry Research Laboratory.

Hernandez-Cuevas, M., Santiago-Garcia, W., Martinez-Antunez, P. Ruiz-Aquino, F. Models of dominant height growth and site indexes for *Pinus ayacahuite* Ehren. AgroSciences, 52(3).

Husch, B., Beers, T., Kershaw, J. Forest Mensuration. 4th edition. John Wiley & Sons. 2003.

Keyser, Chad E., comp. 2008 (revised October 2, 2019). Pacific Northwest Coast (PN) Variant Overview – Forest Vegetation Simulator. Internal Rep. Fort Collins, CO: U. S. Department of Agriculture, Forest Service, Forest Management Service Center. 68p.

Lee, YJ. 1971. Predicting mortality for even-aged stands of lodgepole pine. *Forestry Chronicle.* 47: 29-32

MacKinney, A. L., F. X. Schumacher, and L. F. Chaiken. Construction of yield tables for nonnormal loblolly pine stands. Jour. Agric. Research 54:531-545. 1937.

Monserud, RA. Sterba, H., 1999. Modeling individual tree mortality for Austrian forest species. *Forest Ecology and Management.* 113: 109-123

Reineke, LH. 1933. Perfecting a stand-density index for even-aged forest. *Journal of Agricultural Research.* 46:627-638

Seidl, R., and W. Rammer. 2017. Climate change amplifies the interactions between wind and bark beetle disturbances in forest landscapes. *Landscape Ecology* **32**:1485–1498.

Sit. V., Poulin-Costello, M. 1994. Catalogue of Curves for Curve Fitting. Ministry of Forests Research Program. 42-47.

Somers, Greg L. Oderwald, Richard G., Harms, William R., Langdon, O. Gordon. Forest Science, Volume 26, Issue 2, June 1980, Pages 291–300,

Tappeiner J., Maguire, D., Harrington T., Bailey, J. Silviculture and Ecology of Western U.S. Forests. 2cd edition. Oregon State University Press. 2015.

Temperli, C., H. Bugmann, and C. Elkin. 2013a. Cross-scale interactions among bark beetles, climate change, and wind disturbances: a landscape modeling approach. *Ecological Monographs* 83:383–402.

Vanclay, J. Modelling Forest growth and yield: Applications to Mixed Tropical Forests. Cab International. 1994.

Wykoff, W.R., Crookston, N.L., Stage, A.R., 1982. User's guide to the Stand Prognosis Model. General Technical Report INT-133. USDA Forest Service, North Central Forest Experimental Station, St. Paul, MN. 184-191.

Weiskittel, A., Hann, D., Kershaw J., Vanclay, J., 2011. Forest Growth and Yield Modeling.

Yoda, K. Kira, T., Ogawa, H., Hozumi, K., 1963. Self-thinning in overcrowded pure stands under cultivated and natural conditions. *Journal of Biology, Osaka City University.* 14: 107-129.

Zhang, X., Cao, Q., Duan, A., Zhang, J. 2017. Modeling tree mortality in relation to climate, initial planting density, and competition in Chinese fir plantations using a Bayesian logistic multilevel method. NRC Research Press.

A.2 Supplementary Material

	200 TPA Starting Density					
Age	*PLDEN*	*ELEV*	*SI30*	*AA*	*AB*	*AC*
10	*190.375*	*783.125*	*80.375*	*1518.25*	*168.75*	*3773.5*
15	*203.875*	*792.875*	*82.25*	*7661.875*	*1283.625*	*-*
20	*199.7143*	*719.2857*	*80.85714*	*15152.14*	*3875.714*	*10033.29*
25	*199.7143*	*1033.25*	*79.5*	*23034.5*	*7110.75*	*13379.25*

	440 TPA Starting Density					
10	*412.5*	*783.125*	*80.25*	*7401.25*	*431*	*9026.375*
15	*421*	*886.6667*	*77*	*24304*	*2221.667*	*14457*
20	*406.8333*	*772.5*	*78.83333*	*45518.83*	*5996.667*	*20905.33*
25	*422.25*	*1033.25*	*78.25*	*58697.5*	*9152.5*	*26818.25*

	680 TPA Starting Density					
10	*599.7143*	*781.5714*	*77*	*15543.29*	*599.4286*	*13139.29*
15	*668*	*877.5*	*82*	*60462.5*	*3543.5*	*26137.25*
20	*610.8*	*768.2*	*73.6*	*79205*	*6305.8*	*28747.6*
25	*630.3333*	*1011*	*78*	*107276*	*11053.33*	*40909.67*

Table A.2. Average values for model predictors for all plots at specific ages used in model predictions. If data was not collected at a specific year the slope of the measurement year before and after was used to calculate at the unknown year.

A.3 Acknowledgements

I would like to thank my committee chair Eric Turnblom for giving me this opportunity, endless guidance and knowledge over the past two years. Thank you to my committee members Patrick Clune and Gregory Ettl for their contributions throughout this entire process, and a special thanks to Gregory Ettl for his encouragement and time starting my career in forestry as an undergraduate. Thanks to Robert Swan for showing me the ropes throughout graduate school and keeping me level-headed throughout very unforeseen circumstances. Thanks to Jason Cross for not only his assistance with the model development, but with editing, stepping up and being a mentor when it was needed and keeping me on track to finish. Thanks to the professors and staff of the School of Environmental Science and Forestry for their teaching and guidance over the years. Thank you to the cooperative members for providing the land and trees that this model was developed from and for some of the funding to get me through the program. Finally, a thanks to my roommates and parents for keeping me sane and always on track and supported even in the most stressful of days.