Evaluation and Modeling of One and Multi-Point Douglas-fir Site Indices Using Machine Learning in Northern California

A Thesis
Presented in Partial Fulfillment of the Requirements for the
Degree of Master of Science
with a
Major in Natural Resources
in the
College of Graduate Studies
University of Idaho
by
Logan J. Wimme

Approved by:
Major Professor: Mark Kimsey, Ph.D.
Committee Members: Ann Abbott, Ph.D.; Paul Gessler, Ph.D.
Department Administrator: Charles Goebel, Ph.D.

May 2022

Abstract

Accurately quantifying forest productivity is a vital endeavor for modern forest managers. In north central California, one-point site index equations created from stem analysis data currently serve as the most reliable means to estimate forest productivity. Although generally sufficient, current models may inherently introduce error by failing to acknowledge how site conditions differentially impact growth rates and that growth rates fluctuate as trees mature. Therefore, alternative approaches that implicitly incorporate site growth factors may be necessary to quantify the true productive potential of forested landscapes in this region.

We selected 162 Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco var. menziesii) trees for destructive sampling via segmented stem analysis from a replicated orthogonal sampling matrix of known environmental growth factors. Predicted growth rates for stem segments were calculated using a locally established traditional one-point Douglas-fir site index equation that utilizes breast-height age and total height. These predicted growth rates were then compared to observed growth rates obtained using a two-point site index approach (i.e., age by log segment length). Results indicated no significant differences between observed and predicted growth rates for breast-height to 20 and breast-height to 30 m segments. However, significant underpredictions were identified for a majority of segments between breast-height and 30 m. Results suggest the effectiveness and utility of one and two-point site index approaches is highly dependent on past management practices (available site trees), future silvicultural objectives (short vs long-term rotation lengths), and a need to accurately predict temporal growth rates (carbon accumulation).

To meet the demand for more reliable one and multi-point productivity estimates, extreme gradient boosting (XGBoost) machine learning models using climatic, edaphic, and topographic predictors were tested to evaluate prediction accuracies of Krumland and Eng (2005) site index and 10-meter site index (10MSI) – two site index approaches commonly used to calibrate regional growth and yield models. Multiple XGBoost models were created and compared for each site index method. The lowest 10-fold cross-validated RMSE value was used to select final models for each site index method. Final machine learning models were used to generate 0.4-hectare resolution raster layers of productivity for the study area.

Acknowledgments

I want to thank Dr. Mark Kimsey, my major professor, for his constant support, guidance, and mentorship throughout all phases of this research project. I also thank Ann Abbott for her contributions to my education and statistical advice regarding project analysis. I also want to thank Paul Gessler for his geospatial expertise and continuous words of encouragement. My development as a student and future professional has been highly impacted by you all in a valuable way, and for that I am ever grateful. Additionally, I would like to thank the faculty and staff of the Department of Forest, Rangeland, and Fire Sciences at the University of Idaho for a genuinely positive and productive experience as a graduate student.

I also thank NewForests and their staff for providing the necessary funds and assistance for this study. Lastly, I would like to thank the following individuals who were instrumentally involved with the execution of this research project: Elijah Allensworth, James Bullen, Weikko Jaross, Jacob Strunk, James Arney, Josh Gomes, Mark Slay, Alexandrea Meacham, and Emily Elfering. Completion would not have been possible without your dedicated cooperation.

Dedication

This thesis is dedicated to my parents, Kris and Lisa. You always inspired me to take life further. I thank you for your endless prayers, love, and support.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Although a common and well-researched topic, forest productivity remains an important concept in the minds and actions of modern forest managers (Carmean 1954; Grier et al. 1989; Bontemps and Bouriaud 2014). In basic terms, forest productivity can be defined as a forest's potential to produce aboveground wood volume (Skovsgaard and Vanclay 2008). Estimates of forest productivity as a function of influential site growth factors (i.e., temperature, moisture, nutrition) allow us to predict the rate at which trees can grow at a given spatial location (DeYoung 2016).

Accurate and reliable estimates of productivity are vital for predicting economic returns and maintaining sustainable multi-use or intensively managed forest ecosystems (Dolph 1988; Skovsgaard and Vanclay 2008; Weiskittel et al. 2011). Landowners, managers, and investors are strongly interested in determining future growth and value of timberlands for real estate, atmospheric carbon sequestration, taxation, and resource sustainability purposes, and they realize the consequences of inaccurate productivity estimates (Huston and Marland 2003; Seagle 2008; Newell and Eves 2009). Because productivity strongly impacts projected biomass estimates and future value of timberlands, it is a dictating factor in modern carbon markets and the acquisition and sale of property (Seagle 2008; Newell and Eves 2009). This plays a large role in today's forestry community where timber investment management organizations (TIMOs) and real estate investment trusts (REITs) are responsible for a large portion of industrial timberlands in the United States (Fernholz 2007; Newell and Eves 2009). Industrial landowners are also particularly interested in productivity estimates because they are used in some state systems to adjust property tax rates (Klemperer 1976). Additionally, a forest's productive potential is significant in terms of exploring management options and alternative silvicultural practices (Vanclay 1994; Devaranavadgi et al. 2013). In many cases, certain wood products can only be extracted from a forest when timber reaches a specified size. With accurate productivity estimates, managers can predict when volume will be ready for harvest and build both harvest and planting schedules accordingly (Latta et al. 2010; Devaranavadgi et al. 2013; Bontemps and Bouriaud 2014). If managers know the productivity of a desired species for a range of site conditions, they also have the luxury of strategically planting sites that will yield the greatest returns (Woolery et al. 2002; Aertsen et al. 2012).

Numerous studies have shown that correlations between productivity and environmental factors may provide the most suitable approach to estimate forest productivity (Brown and Loewenstein 1978; Grier et al. 1989; Kimsey et al. 2008; Aertsen et al. 2012; Bontemps and Bouriaud 2014; Parresol et al. 2017; Hemingway 2020). Environmental factors that are proven to influence site quality and therefore drive forest productivity include climate, edaphic properties, topography, and light (Corona et al. 1998; Skovsgaard and Vanclay 2008; DeYoung 2016). Measurable variables such as temperature, precipitation, soil depth, available water supply, parent material, elevation, slope, aspect, and length of growing season all interact to determine the rates at which trees grow (Grier et al. 1989; DeYoung 2016).

Some early approaches to estimate productivity attempted to quantify these site factors in order to create a balanced sampling design (Skovsgaard and Vanclay 2008). However, the absence of widespread, reliable site condition data and modern geographic information systems (GIS) made it difficult to properly stratify expansive landscapes (Monserud 1984; Hemingway 2020). Proper stratification of the landscape is essential when generating accurate productivity estimates to ensure all micro-site conditions are represented in models.

Early predictions of forest productivity were primarily obtained from site-specific experience tables, which were soon replaced by yield tables and growth models (Skovsgaard and Vanclay 2008). Since then, many variations of growth models have been generated and implemented. Beginning in the early twentieth century, estimates of stand height at a given age, commonly known as site index, has been the standard means of quantifying and forecasting forest productivity (Powers 1972; Hägglund and Lundmark 1977; Brown and Loewenstein 1978; Batho and Garcia 2006; Parresol et al. 2017). Height over age curves now exist for most commercial timber species in the United States (Powers 1972). Site index approaches, where only one height-age pair is needed, are amongst the most widely used methods to quantify site productivity (Batho and Garcia 2006b; Monserud et al. 2006;

Kimsey et al. 2008; Hemingway 2020), and are used within many variants of the Forest Vegetation Simulation (FVS) growth and yield software ecosystem.

Growth of dominant and codominant trees is best represented by a sigmoidal curve because growth rates vary as trees progress through different stages of maturity (Weiner and Thomas 2001). Rates are gradual in early-life stages, increase in middle stages, and plateau in late stages (Weiner and Thomas 2001; Hemingway 2020). Because growth rates are slow and gradual in early and late stages, and trees in early stages are prone to impacts of non-productivity related factors, it makes practical sense to determine site productivity based on the middle stage of a tree's life (Arney et al. 2009; Hemingway 2020). It is in this stage where growth rates are the highest and recruitment of wood fiber is most reflective of the site's potential (Arney et al. 2009; Hemingway 2020).

Thus, to generate more accurate and unbiased productivity estimates, a two-point site index approach may provide the best solution (Arney et al. 2009; Hemingway 2020; Zeide, 1978). The two-point method, first introduced by Zeide (1978), explains how multiple height-age measurement pairs for a selected dominant or codominant site tree can be used to provide more accurate estimates of site productivity (Arney et al. 2009; Hemingway 2020; Zeide, 1978). This method essentially eliminates the need for traditional site index equations (Arney et al. 2009; Hemingway 2020; Zeide, 1978). The 10m Site Index (10MSI) method is an approach that encompasses Zeide's two-point concept and eliminates biases from both early and late stage tree growth (Arney et al. 2009; Hemingway 2020). Two height-age measurements are used; one at 10 meters, and the other at 20 meters (Arney et al. 2009; Hemingway 2020). A growth rate expressed in meters per decade can be extracted from these measurements and used to define site productivity (Arney et al. 2009; Hemingway 2020; Zeide, 1978), and is the backbone of growth and yield modeling in the widely used Forest Projection and Planning System (FPS) software package (Forest Biometrics Research Institute, 2020).

Numerous site index equations and models for Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco var. menziesii) are currently in use across the western united states (Cochran 1979; Monserud 1984; Hann and Scrivani 1987). However, the incentive for refined productivity estimates has created a demand for method evaluation and improved accuracies.

This demand is especially prevalent in portions of California. Throughout the last one hundred years, more than twenty site index models have been implemented for use in the State of California, many of which were adopted from neighboring regions (Krumland and Eng 2005). Many of these models were created by traditional anamorphic guide curve techniques using only one height-age pair from selected sample trees (Bruce 1926; Krumland and Eng 2005). Although simplistic, error and lack of consistency resulting from these approaches suggest that stem analysis methods may serve as a superior alternative (Curtis 1964; Krumland and Eng 2005).

Douglas-fir site index curves created from stem analysis data by Krumland and Eng (2005) appear to be more reliable than predecessors for forested regions of northern California. These base age invariant curves were built specifically for unique and homogenous portions of the state, including our area of interest in north central California (Krumland and Eng 2005). Although typically dependable, these curves utilize the one point method, which by default can confound productivity estimates if site trees are not screened for early growth impacts due to silvicultural treatments, suppression, disturbance, and other early-life environmental impacts (Arney et al. 2009; Hemingway 2020). Due to the importance of site index to regional growth and yield models (FVS, FPS) and for sustainable land use planning, our study was designed to evaluate regional one vs multi-point site index equations for accuracy and bias across ranges of climatic, edaphic, and topographic growth factors using tree stem height-age pairs. Proper stratification of the study area was especially important because these growth factors vary greatly from stand to stand in north central California (Dunning and Reineke 1933; Baker 1944). Our research objectives were to 1) determine if significant differences in segmented tree growth rates existed between the one-point Krumland and Eng (2005) site index approach and the two-point 10MSI site index approach, 2) determine where along the stem significant differences in growth rates occur, 3) quantify significant differences in growth rates, and 4) model Krumland and Eng (2005) site index and 10MSI across the study area using measurable site-specific predictors.

Chapter 2: Evaluation of One and Multi-Point Douglas-fir Site Indices

Abstract

Accurately quantifying forest productivity is a vital endeavor for modern forest managers. In north central California, one-point site index equations created from stem analysis data currently serve as the most reliable means to estimate forest productivity. Although generally sufficient, current equations may inherently introduce error by failing to acknowledge differences in growth rates as trees mature. Therefore, alternative approaches may be necessary to quantify the true productive potential of forested landscapes in this region. We selected 162 Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco var. menziesii) trees for destructive sampling via segmented stem analysis from a replicated orthogonal sampling matrix across a range of known environmental growth factors. Sample trees were segmented and age was recorded at breast-height (1.37 m), 10, 20, and 30 m positions. We compared growth rates predicted from a locally established one-point Douglas-fir site index equation to observed growth rates for 1.37 to 10 m, 1.37 to 20 m, 1.37 to 30 m, 10 to 20 m, 10 to 30 m, and 20 to 30 m stem segments. Results indicated no significant differences between observed and predicted growth rates for breast-height to 20 and breast-height to 30 m segments, indicating Krumland and Eng (2005) site index adequately quantifies tree-length growth and may remain informative for various management practices in this region. However, significant underpredictions were identified for a majority of other log segments, including the 10 to 20 m segment often used in the 10MSI method. On average, Krumland and Eng (2005) site index overpredicted growth rates for breast-height to 10 m segments by 0.28 meters per decade and underpredicted for 10 to 20, 10 to 30, and 20 to 30 m segments by 0.57, 0.47, 0.46 meters per decade respectively. Predicted growth rates for breast-height to 20 and breast height to 30 m segments were consistent with observed growth rates.

These findings suggest a two-point approach may be superior when there is an economic or ecological need to quantify site-specific, temporal growth rates other than at total tree age. It is concluded that practicality and effectiveness of one and multi-point site index approaches in this region is heavily reliant on past silviculture, current and future management objectives, and user standard operating procedures.

Introduction

Although a common and well-researched topic, forest productivity remains an important concept in the minds and actions of modern forest managers (Carmean 1954; Grier et al. 1989; Bontemps and Bouriaud 2014). In basic terms, forest productivity can be defined as a forest's potential to produce aboveground wood volume (Skovsgaard and Vanclay 2008).

Accurate and reliable estimates of productivity are vital for predicting economic returns and maintaining sustainable multi-use or intensively managed forest ecosystems (Dolph 1988; Skovsgaard and Vanclay 2008; Weiskittel et al. 2011). Landowners, managers, and investors are strongly interested in determining future growth and value of timberlands for real estate, atmospheric carbon sequestration, taxation, and resource sustainability purposes, and they realize the consequences of inaccurate productivity estimates (Huston and Marland 2003; Seagle 2008; Newell and Eves 2009). A forest's productive potential is also a significant factor of consideration when exploring management options and prescribing silvicultural treatments (Vanclay 1994; Devaranavadgi et al. 2013). Additionally, productivity estimates are commonly used to predict economical rotation ages, therefore laying the foundation for harvest and planting schedules (Latta et al. 2010; Devaranavadgi et al. 2013; Bontemps and Bouriaud 2014).

Numerous studies have shown that correlations between productivity and environmental factors may provide the most suitable approach to estimate forest productivity (Brown and Loewenstein 1978; Grier et al. 1989; Kimsey et al. 2008; Aertsen et al. 2012; Bontemps and Bouriaud 2014; Parresol et al. 2017; Hemingway 2020). Climate, edaphic properties, topography, and light have all been proven to influence site quality and therefore drive forest productivity (Corona et al. 1998; Skovsgaard and Vanclay 2008; DeYoung 2016). Measurable variables such as temperature, precipitation, soil depth, available water supply, soil parent material, elevation, slope, aspect, and length of growing season all interact to determine the rates at which trees grow (Grier et al. 1989; DeYoung 2016).

Some early approaches to estimate productivity attempted to quantify these site factors in order to create a balanced sampling design (Skovsgaard and Vanclay 2008). However, the absence of widespread, reliable site condition data and modern geographic information systems (GIS) made the effective stratification of expansive landscapes dubious (Monserud 1984; Hemingway 2020). Proper stratification methods are imperative when trying to capture effects of micro-site conditions and generate accurate productivity estimates.

Early predictions of forest productivity were primarily obtained from site-specific experience tables, which were superseded by yield tables and the implementation of growth models (Skovsgaard and Vanclay 2008). Since then, many variations of growth models have been generated and rooted in forest operations. Beginning in the early twentieth century, estimates of stand height at a given age, commonly known as site index, has been the standard means of quantifying and forecasting forest productivity (Powers 1972; Hägglund and Lundmark 1977; Brown and Loewenstein 1978; Batho and Garcia 2006; Parresol et al. 2017). Height over age curves currently exist for most commercial timber species in the United States (Powers 1972). One point site index approaches, where only one height-age pair is needed to quantify site productivity, are amongst the most widely used methods (Batho and Garcia 2006b; Monserud et al. 2006; Kimsey et al. 2008; Hemingway 2020). One point approaches are favored for their simplicity, and they are heavily used within many variants of the Forest Vegetation Simulation (FVS) growth and yield software ecosystem.

Growth of dominant and codominant tree growth is best represented by a sigmoidal curve because growth rates fluctuate as trees mature (Weiner and Thomas 2001). Rates are gradual in early-life stages, increase in middle stages, and plateau in late stages (Weiner and Thomas 2001; Hemingway 2020). Because growth rates are slow and gradual in early and late stages, and trees in early stages are prone to impacts of non-productivity related factors, determining site productivity according to the middle stage of a tree’s life makes practical sense (Arney et al. 2009; Hemingway 2020). It is in this stage where growth rates are most reflective of the site’s potential (Arney et al. 2009; Hemingway 2020).

Thus, to generate more accurate and unbiased productivity estimates, a two-point site index approach may provide the best solution (Arney et al. 2009; Hemingway 2020; Zeide, 1978). The two-point method, first introduced by Zeide (1978), explains how multiple height-age measurement pairs for a selected dominant or codominant site tree can be used to calculate true growth rates and provide more accurate productivity estimates (Arney et al.

2009; Hemingway 2020; Zeide, 1978). Two ring count measurements and the length of stem between measurements can be used to express growth rates in meters per decade (Arney et al. 2009; Hemingway 2020; Zeide, 1978). This essentially eliminates the need for traditional site index equations (Arney et al. 2009; Hemingway 2020; Zeide, 1978), which is why some growth and yield modeling applications, like the Forest Projection and Planning System (FPS), favor this approach.

Numerous site index equations and models for Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco var. menziesii) are currently used across the western United States (Cochran 1979; Monserud 1984; Hann and Scrivani 1987). However, the incentive for refined productivity estimates has created a demand for method evaluation and improved accuracies. This demand is especially prevalent in portions of California. More than twenty site index models have been implemented for use in the State, many of which were adopted from neighboring regions (Krumland and Eng 2005). Many of these models were created by traditional anamorphic guide curve techniques using only one height-age pair from selected sample trees (Bruce 1926; Krumland and Eng 2005). Although simplistic, error and inconsistencies suggest that stem analysis methods serve as a superior alternative (Curtis 1964; Krumland and Eng 2005).

Douglas-fir site index curves created from stem analysis data by Krumland and Eng (2005) appear to be more reliable than predecessors for forested regions of northern California. These base age invariant curves were built specifically for unique and homogenous portions of the state, including our area of interest in north central California (Krumland and Eng 2005). Although typically dependable, these curves utilize the one point method, which by default can confound productivity estimates if site trees are not screened for early growth impacts due to silvicultural treatments, suppression, disturbance, and other early-life environmental impacts (Arney et al. 2009; Hemingway 2020).

Due to the importance of site index to regional growth and yield models (FVS, FPS, etc.) and for sustainable land use planning, our study was designed to evaluate regional one vs multi-point site index equations for accuracy and bias across ranges of climatic, edaphic, and topographic growth factors using tree stem height-age pairs. Proper stratification of the study area was especially important as these growth factors vary greatly from stand to stand in north central California (Dunning and Reineke 1933; Baker 1944). The objectives of this study were to 1) determine if significant differences in segmented tree growth rates existed between the one-point Krumland and Eng (2005) site index approach and the two-point site index approach, 2) determine where along the stem significant differences in growth rates occur, and 3) quantify significant differences in growth rates where they do occur.

Methods and Materials

Study Area

The study area was a 66,283-hectare land ownership that covered parts of Shasta, Trinity, and Siskiyou counties in north central California, USA (Fig. 2.1). Site conditions varied greatly across the study area which resulted in wide ranges of values for measurable variables. Elevations ranged from 435 m up to 2198 m with an average of 1237 m. Mean annual precipitation ranged from 586 mm to 1939 mm with an average of 1409 mm. Regional geologic soil parent materials included ultramafic rocks and serpentine soils in the west, to volcanics located around Mt Shasta. Depths to restrictive layers ranged from 0 cm (bedrock) to 251 cm with an average of 137 cm. Although Douglas-fir was a large component of most forested stands, other common species present on the landscape included white fir (*Abies concolor* (Gord.) Lemm var. lowiana), ponderosa pine (*Pinus ponderosa* Dougl. ex Laws. var. ponderosa), sugar pine (*Pinus lambertiana*), and California incense cedar (*Calocendrus decurrens*).

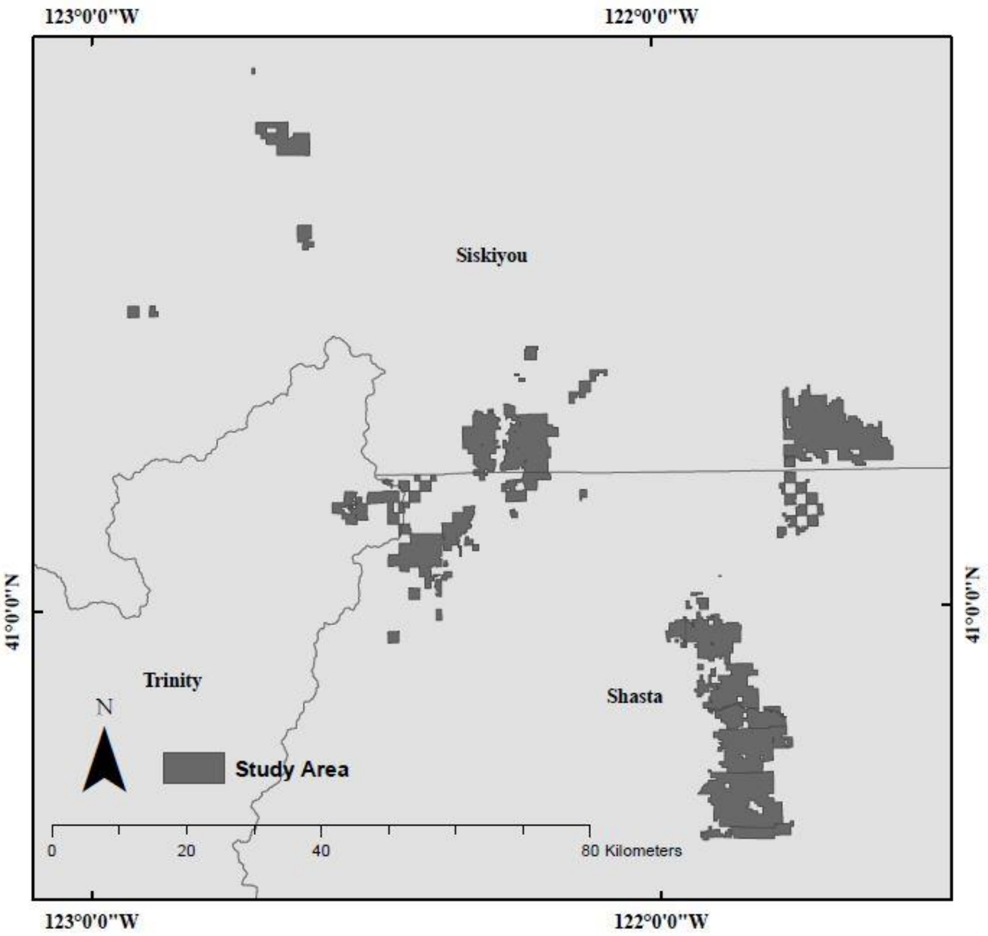

Figure 2.1. Study Area.

Stratification and Sample Selection

The study area was stratified by known tree growth factors (temperature, moisture, soil environments) to ensure samples were collected across the range of varying site conditions (Hemingway 2020). A 0.4-hectare point grid was laid across the entire landscape and attributes of mean annual precipitation (MAP), degree days above 5°C (DD5), solar radiation (RAD), depth to restrictive layer (DEP2RESLYR), available water supply from 0 to 100 cm (AWS100), geologic soil parent material – a proxy for soil nutrition and development (SPM), and elevation (EL) were obtained for each point. The interaction of DD5 and scaled RAD values were used to derive a surface heat load (HL) value. DEP2RESLYR, AWS100, and SPM values were used to generate a single soil metric termed soil quality index (SQI).

MAP and DD5 values were obtained from ClimateNA 1961-1990 30-yr normals (Wang et al. 2016). RAD values were computed using ESRI’s ArcMap 10.7.1 Points Solar Radiation tool. Soil data used to generate SQI values were obtained from Gridded National Soil Survey Geographic Database (gNATSGO) soil maps (Natural Resources Conservation Service, 2020). EL values were obtained from a 30-m United States Geological Survey (USGS) digital elevation model (DEM). All soil and elevation data were extracted to point locations using GIS extraction tools in ArcMap 10.7.1. Recent timber cruise data for the study area was used to remove points where species other than Douglas-fir were the dominant stand component.

An orthogonal sampling matrix using MAP, HL, and SQI was used to create 27 strata (Fig. 2.2; Arney et al. 2009; Hemingway and Kimsey 2020). Three strata bins for the continuous variables MAP and HL were defined as three equally proportionate bins by setting breaks at the mean of each variable ± ½ standard deviation. Potential sampling points in the lowest bin (<½ standard deviation) were given a value of 1, points with values in the middle bin (± ½ standard deviation) were given a value of 2, and points in the highest bin (>½ standard deviation) were given a value of 3.

Previously measured Douglas-fir site index values from regional stand inventories were used as proxies for SPM quality (i.e., a means to develop a continuous proxy variable for SPM strata). Average site index values across each SPM were then used to rank SPMs from 1 to 15 with 1 being the least productive and 15 being the most productive. SPM strata breaks were created using the mean ± ½ standard deviation of averaged site index values. Potential sampling points with SPMs in the lowest site index bin were assigned a value of 1, those with parent materials in the middle bin were assigned a value of 2, and those in the highest bin were assigned a value of 3.

For DEP2RESLYR and AWS100, point values were similarly divided into three proportionate bins by setting breaks at the mean ± ½ standard deviation. Points with values in lowest bin for each variable were assigned a 1, points with values in the middle bin for each variable were assigned a 2, and points with values in the highest bin for each variable were assigned a 3. Bin numbers for DEP2RESLYR, AWS100, and SPM were then added together and divided by three to obtain an average for each point. Averaged values for points were then placed into three equally proportionate bins by setting breaks at the mean ± ½ standard deviation. Points that fell into the lowest bin were given a SQI value of 1, points that fell into the middle bin were given a SQI value of 2, and points that fell into the highest bin were given a SQI value of 3. The workflow for obtaining SQI values is shown in Figure 2.3.

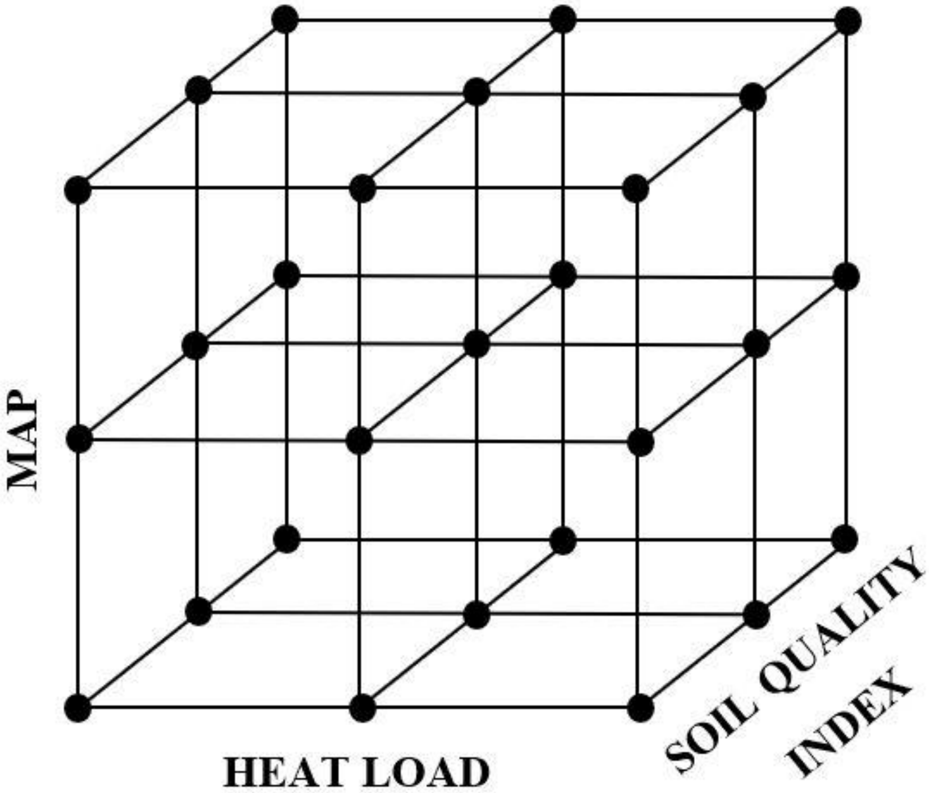

Figure 2.2. Orthogonal sampling matrix.

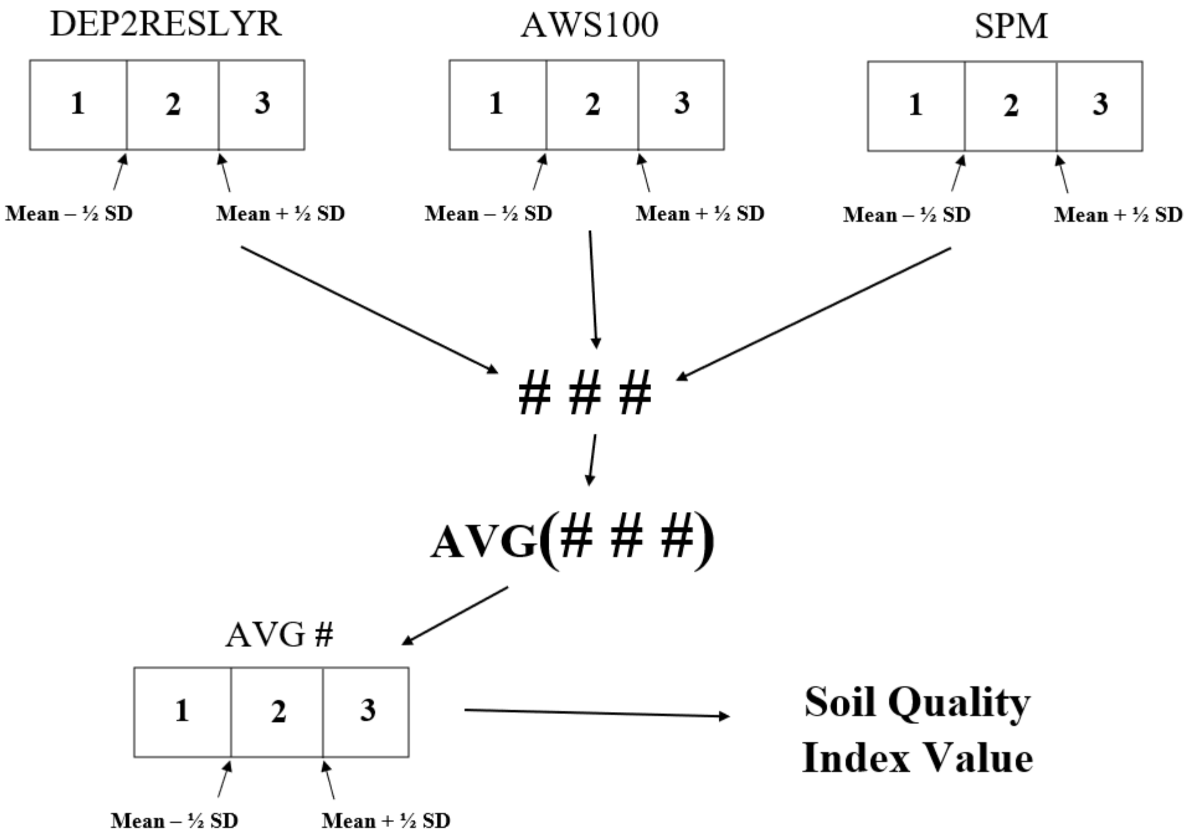

Figure 2.3. Workflow for obtaining soil quality index values.

The 27 unique strata defined by MAP, HL, and SQI were then replicated for both low elevation regions (<1200 m) and high elevation regions (≥1200 m) for a total of 54 strata. This stratification approach provided us the ability to sample across the range of MAP, HL, SQI, and EL values of the study area in order to capture variation in regional and stand-level growing conditions.

Sample locations from the 0.4 hectare study area grid were randomly selected across the 54 strata. Visual inspections of each selected sample location were conducted in the summer of 2020 to verify 1) the presence of suitable Douglas-fir site trees and 2) accessibility. To be considered a suitable site tree, a tree needed to be consistent with the following: 1) dominant or codominant canopy position, 2) at least 20 m tall, 3) evidence of consistent growth from an early age – indicated by evenly spaced branch whorls along the tree bole, 4) no evident defect or sign of illness, 5) not growing in an advantageous location relative to other trees in the stand, and 6) safely accessible to a tree feller. If a sample location was deemed insufficient in regard to suitable site trees or accessibility to such trees, a new location from the same strata was randomly selected for inspection. The number of selected sites per strata ranged from 0 to 3. Several strata were not represented due to a limited areal presence within the study area and/or a lack of sufficient site trees. Additionally, several strata were represented more than once due to an extensive areal presence within the study area. In total, 81 sample locations were selected across the 54 strata, representing the diverse landscape of the region (Fig. 2.4, Tables 2.1 and 2.2).

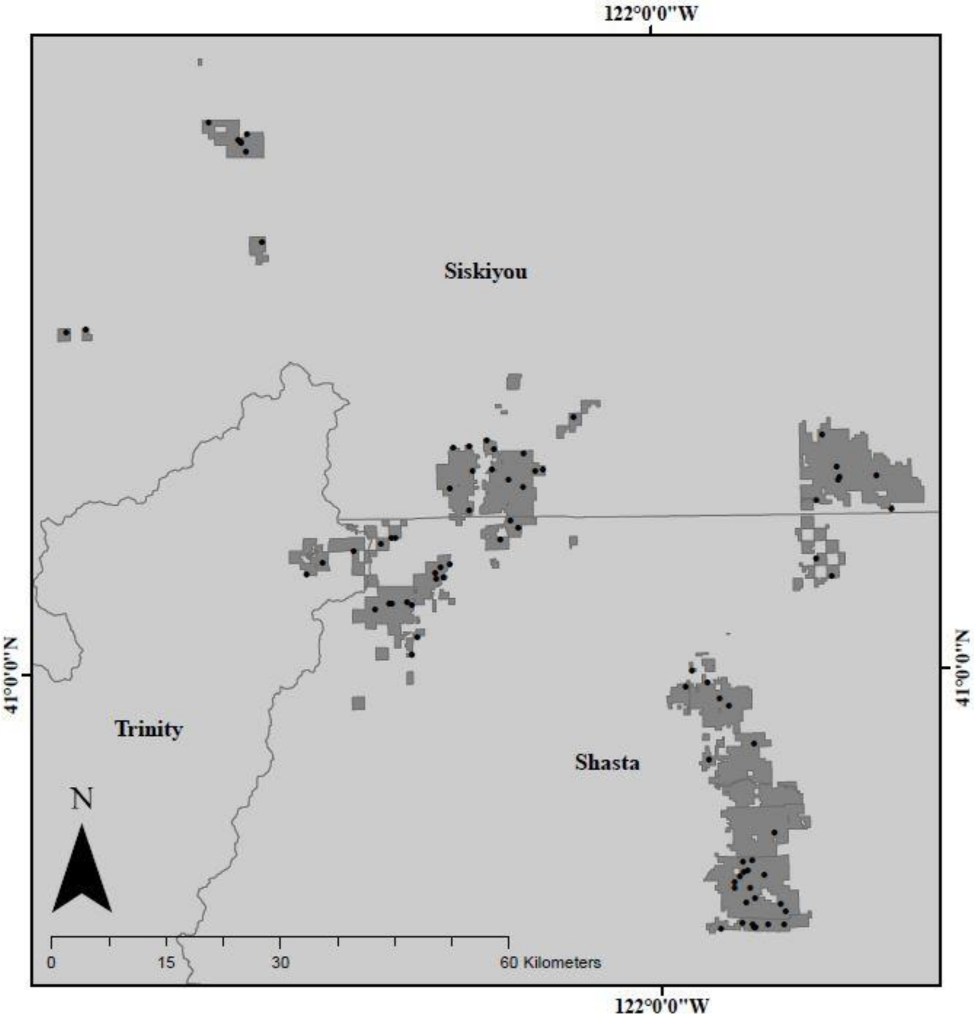

Figure 2.4. Spatial distribution of selected sample locations.

Table 2.1. Summary of site conditions for selected sample locations.

	Range	*Mean*	*SD*
EL[1] (m)	553 - 1596	1173	229
MAP[2] (mm)	637 - 1906	1396.4	359.9
DD5[3]	1878 - 3414	2583.2	356.6
RAD[4] (WH/m^2)	644035 - 1547840	1279950	209694
AWS100[5] (cm)	3.5 - 38.21	13.1	8.3
DEP2RESLYR[6] (cm)	23 - 201	129.8	50.8

[1] Elevation

[2] Mean annual precipitation

[3] Degree days greater than or equal to 5 °C

[4] Solar radiation

[5] Available water supply from ground surface to 100 cm

[6] Depth to restrictive layer

Table 2.2. Representation of geologic soil parent material types (SPM) amongst selected sample locations.

SPM	*% of Sample Locations*
Alluvium	1.2
Colluvium	2.5
Serpentine	4.9
Ash over volcanics	3.7
Volcanics	30.9
Tephra	11.1
Ultramafic	9.9
Igneous (unknown)	1.2
Igneous (extrusive)	3.7
Igneous (intrusive)	1.2
Metamorphic	28.4
Glacial	1.2

Sampling

Stem analysis sampling of Douglas-fir was completed during the summer of 2021. Felled tree measurements for two trees with similar 10 to 20 m growth rates were collected at each sample location. A third tree was selected and felled if one of the first two trees expressed a 10 to 20 m growth rate that fell beyond plus or minus 15 percent of the mean 10 to 20 m growth rate of the first two trees. Only data from the two most similar trees at each location were retained. This ensured that collected samples were consistent in representing each site. Once felled, cross sections were obtained at 0.3 m (stump), 1.37 m (breast-height), 10 m, 20 m, and if possible 30 m. Ring count (age) was recorded for each cross section. In total, data from 162 trees were retained (Table 2.3).

Table 2.3. Summary of sample tree measurements and ages.

	Range	*Mean*	*SD*
Total Height (m)	21.2 – 40.3	30.8	14.2
DBH (cm)	35.6 – 88.9	54.9	9.2
Age at 0.3 m	33 - 168	84	22
Age at 1.37 m	31 - 151	77	21
Age at 10 m	21 - 111	53	17
Age at 20 m	5 - 83	32	13
Age at 30 m	1 - 33	13	8

Calculation of observed growth rates

Cross section age data was used to calculate observed growth rates for segments within each sample tree.

BH (breast-height) to 10 m growth rates were calculated with

$$m/d = \frac{86}{age_{BH} - age_{10}} \quad \text{[Eq. 2.1]}$$

where *m/d* is the observed meters per decade growth rate, 86 is the product of multiplying the length from BH to 10 m (8.6 m) by 10 to convert into meters per decade, age_{BH} is the observed age at breast height, and age_{10} is the observed age at 10 m.

BH to 20 m growth rates were calculated with

$$m/d = \frac{186}{age_{BH} - age_{20}} \qquad \text{[Eq. 2.2]}$$

where *m/d* is the observed meters per decade growth rate, 186 is the product of multiplying the length from BH to 20 m (18.6 m) by 10 to convert into meters per decade, age_{BH} is the observed age at breast height, and age_{20} is the observed age at 20 m.

BH to 30 m growth rates were calculated with

$$m/d = \frac{286}{age_{BH} - age_{30}} \qquad \text{[Eq. 2.3]}$$

where *m/d* is the observed meters per decade growth rate, 286 is the product of multiplying the length from BH to 30 m (28.6 m) by 10 to convert into meters per decade, age_{BH} is the observed age at breast height, and age_{30} is the observed age at 30 m.

10 to 30 m growth rates were calculated with

$$m/d = \frac{200}{age_{10} - age_{30}} \qquad \text{[Eq. 2.4]}$$

where *m/d* is the observed meters per decade growth rate, 200 is the product of multiplying the length from 10 to 30 m (20 m) by 10 to convert into meters per decade, age_{10} is the observed age at 10 m, and age_{30} is the observed age at 30 m.

10 to 20 and 20 to 30 m growth rates were calculated with

$$m/d = \frac{100}{age_{k} - age_{k+10}} \qquad \text{[Eq. 2.5]}$$

where *m/d* is the observed meters per decade growth rate, 100 is the product of multiplying the length from k to k+10 m (10 m) by 10 to convert into meters per decade, age_{k} is the observed age at k m, and age_{k+10} is the observed age at k+10 m.

Krumland and Eng site index for sampled trees

Krumland and Eng site index was calculated for each sample tree using the DFI_CR2_MMC model and a base age of 50 (Krumland and Eng 2005). This model was built for interior Douglas-fir in the main mixed-conifer zone of California. The DFI_CR2_MMC model equation is written as

$$H = 4.5 + (H_0 - 4.5) \left\{ \frac{[1-exp(-0.01564*A)]}{[1-exp(-0.01564*A_0)]} \right\}^{(-6.260+38.98/R_0)} \qquad \text{[Eq. 2.6]}$$

where H is site index, H_0 is observed total tree height, A is a specified base age, A_0 is observed age at BH, and R_0 is an unobservable site productivity or growth intensity variable. R_0 can be calculated using

$$R_0 = \frac{\left(L_0-(-6.260*Y_0)\right)+\sqrt{\left(L_0-(-6.260*Y_0)\right)^2-4(38.98)Y_0}}{2} \qquad \text{[Eq. 2.7]}$$

letting

$$L_0 = ln(H_0 - 4.5) \qquad \text{[Eq. 2.8]}$$

$$Y_0 = ln\left(1 - exp(-0.01564 * A_0)\right) \qquad \text{[Eq. 2.9]}$$

where H_0 is observed total tree height and A_0 is observed age at BH.

Krumland and Eng site index predicted BH ages and growth rates

We used the CR2 model form, observed Krumland and Eng (2005) base age 50 site index values, and set heights of 10 m, 20 m, and 30 m to calculate predicted BH ages. We substituted site index values for H and values of 10 m, 20 m, and 30 m for H_0 to solve for corresponding BH ages (A_0).

Height values of BH, 10, 20, and 30 m along with corresponding predicted BH ages were used to calculate predicted growth rates for segments within each sample tree.

Predicted BH to 10 m growth rates were calculated with

$$m/d = \frac{86}{BHA_{10}} \qquad \text{[Eq. 2.10]}$$

where *m/d* is the observed meters per decade growth rate, 86 is the product of multiplying the length from BH to 10 m (8.6 m) by 10 to convert into meters per decade, and BHA_{10} is the predicted BH age when total tree height is 10 m.

Predicted BH to 20 m growth rates were calculated with

$$m/d = \frac{186}{BHA_{20}} \qquad \text{[Eq. 2.11]}$$

where *m/d* is the observed meters per decade growth rate, 186 is the product of multiplying the length from BH to 20 m (18.6 m) by 10 to convert into meters per decade, and BHA_{20} is the predicted BH age when total tree height is 20 m.

Predicted BH to 30 m growth rates were calculated with

$$m/d = \frac{286}{BHA_{30}} \qquad \text{[Eq. 2.12]}$$

where *m/d* is the observed meters per decade growth rate, 286 is the product of multiplying the length from BH to 30 m (28.6 m) by 10 to convert into meters per decade, and BHA_{30} is the predicted BH age when total tree height is 30 m.

Predicted 10 to 30 m growth rates were calculated with

$$m/d = \frac{200}{BHA_{30} - BHA_{10}} \qquad \text{[Eq. 2.13]}$$

where *m/d* is the observed meters per decade growth rate, 200 is the product of multiplying the length from 10 to 30 m (20 m) by 10 to convert into meters per decade, BHA_{30} is the predicted BH age when total tree height is 30 m, and BHA_{10} is the predicted BH age when total tree height is 10 m.

Predicted 10 to 20 and 20 to 30 m growth rates were calculated with

$$m/d = \frac{100}{BHA_{k+10} - BHA_{k}} \qquad \text{[Eq. 2.14]}$$

where *m/d* is the observed meters per decade growth rate, 100 is the product of multiplying the length from k to k+10 m (10 m) by 10 to convert into meters per decade, BHA_{k+10} is the predicted BH age when total tree height is k+10 m, and BHA_{k} is the predicted BH age when total tree height is k m.

Comparison of observed and predicted growth rates

Analysis of covariance (ANCOVA) was used to test for significant differences between observed and predicted growth rates at a 95% confidence level in R (R Core Team, 2021). The aov() function was used to fit a relationship between segment growth rates and Krumland and Eng (2005) site index with a covariate identifying growth rates as either observed or predicted. Significance of covariate P-values were determined by comparing them to an alpha value of 0.05. For significant p-values, the TukeyHSD() function in R was used to quantify differences (R Core Team, 2021). Additionally, standard linear regression was used to observe the relationship between observed and predicted growth rates for tree segments. A 1:1 line was used to visualize the direction and severity of deviance, if any, between growth rates.

Results

Observed BH to 10 m growth rates varied from 1.7 to 10.8 meters per decade with a mean value of 4.1 and a standard deviation of 1.6. Krumland and Eng (2005) predicted BH to 10 m growth rates varied from 1.8 to 12.3 meters per decade with a mean value of 4.4 and a standard deviation of 1.7. Predicted BH to 10 m growth rates averaged 0.28 ± 0.006 meters per decade greater than observed BH to 10 m growth rates. This difference was significant at alpha=0.05 ($p<0.01$) (Fig. 2.5a).

Observed BH to 20 m growth rates varied from 2.2 to 9.8 meters per decade with a mean value of 4.5 and a standard deviation of 1.4. Krumland and Eng (2005) predicted BH to 20 m growth rates varied from 2.1 to 10.6 meters per decade with a mean value of 4.5 and a standard deviation of 1.4. A significant difference between observed BH to 20 m and Krumland and Eng (2005) BH to 20 m growth rates was not detected (Fig. 2.5b).

Observed BH to 30 m growth rates varied from 2.2 to 7.7 meters per decade with a mean value of 4.8 and a standard deviation of 1.1. Krumland and Eng (2005) predicted BH to 30 m growth rates varied from 2.4 to 7.8 meters per decade with a mean value of 4.7 and a standard deviation of 1.0. A significant difference between observed BH to 30 m and Krumland and Eng (2005) BH to 30 m growth rates was not detected (Fig 2.5c).

Observed 10 to 20 m growth rates varied from 1.7 to 9.1 meters per decade with a mean value of 5.2 and a standard deviation of 1.5. Krumland and Eng (2005) predicted 10 to 20 m growth rates varied from 2.3 to 9.5 meters per decade with a mean value of 4.7 and a standard deviation of 1.3. Predicted 10 to 20 m growth rates averaged 0.57 ± 0.004 meters per decade less than observed 10 to 20 m growth rates. This difference was significant at alpha=0.05 (p<0.001) (Fig 2.5d).

Observed 10 to 30 m growth rates varied from 2.8 to 8.7 meters per decade with a mean value of 5.1 and a standard deviation of 1.3. Krumland and Eng (2005) predicted 10 to 30 m growth rates varied from 2.4 to 7.2 meters per decade with a mean value of 4.6 and a standard deviation of 0.9. Predicted 10 to 30 m growth rates averaged 0.47 ± 0.004 meters per decade less than observed 10 to 30 m growth rates. This difference was significant at alpha=0.05 (p<0.001) (Fig. 2.5e).

Observed 20 to 30 m growth rates varied from 1.7 to 10 meters per decade with a mean value of 4.6 and a standard deviation of 1.4. Krumland and Eng (2005) predicted 20 to 30 m growth rates varied from 2.1 to 6.5 meters per decade with a mean value of 4.2 and a standard deviation of 0.9. Predicted 20 to 30 m growth rates averaged 0.46 ± 0.005 meters per decade less than observed 20 to 30 m growth rates. This difference was significant at alpha=0.05 (p<0.001) (Fig. 2.5f).

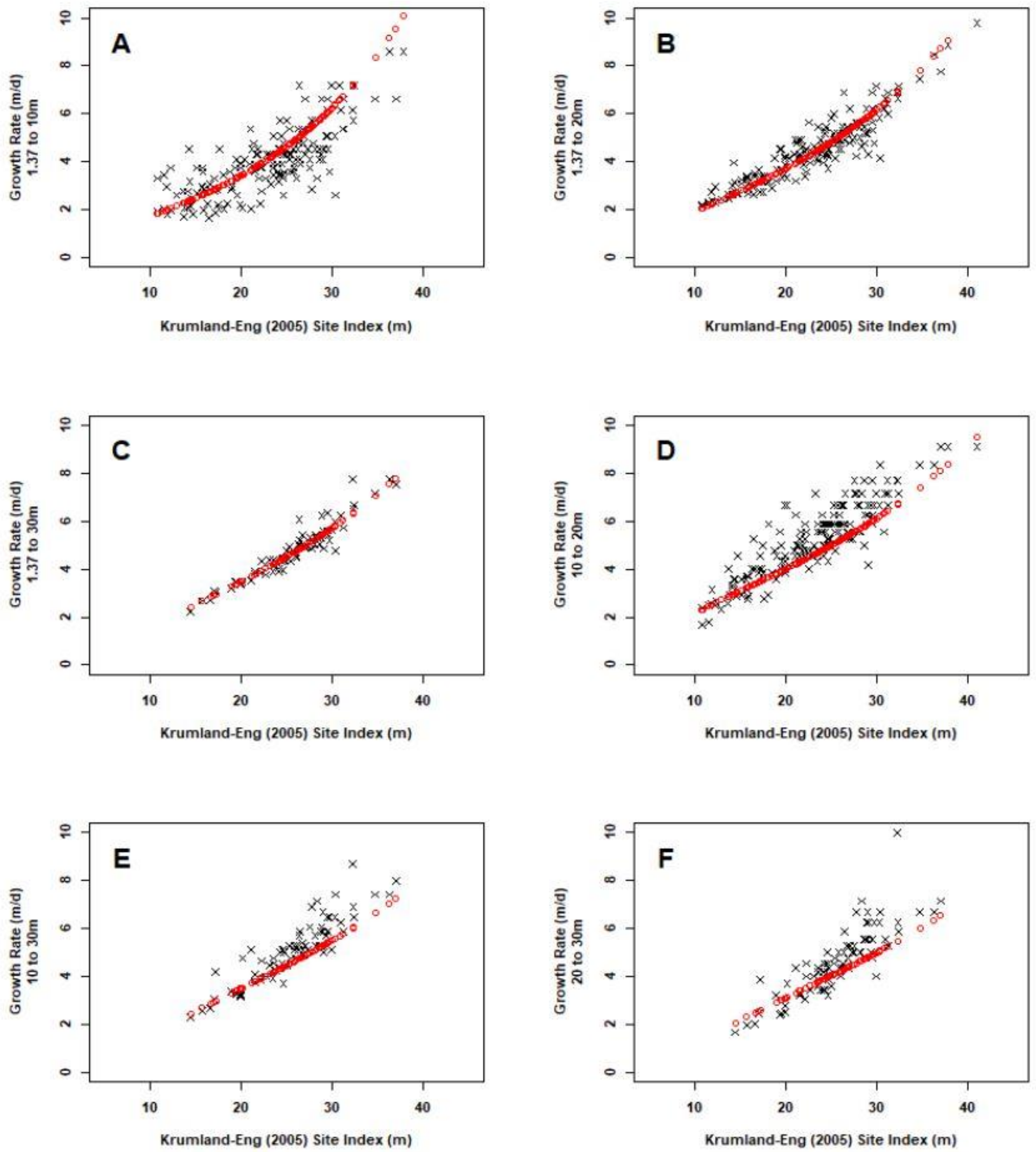

Figure 2.5. Observed and Krumland and Eng (2005) predicted meters per decade growth rates across Krumland and Eng (2005) site index values.

Comparison with a 1:1 line indicated that Krumland and Eng (2005) BH to 20 m and BH to 30 m growth rate predictions were generally consistent with observed values. However, clear inaccuracies existed for all other tree segments (Fig. 2.6). The majority of Krumland and Eng (2005) predicted BH to 10 m growth rates were overpredicted, especially for slower growing trees (Fig. 2.6a). Additionally, the majority of Krumland and Eng (2005) predicted 10 to 20 m, 10 to 30 m, and 20 to 30 m growth rates were underpredicted, especially for faster growing trees (Fig. 2.6d-f).

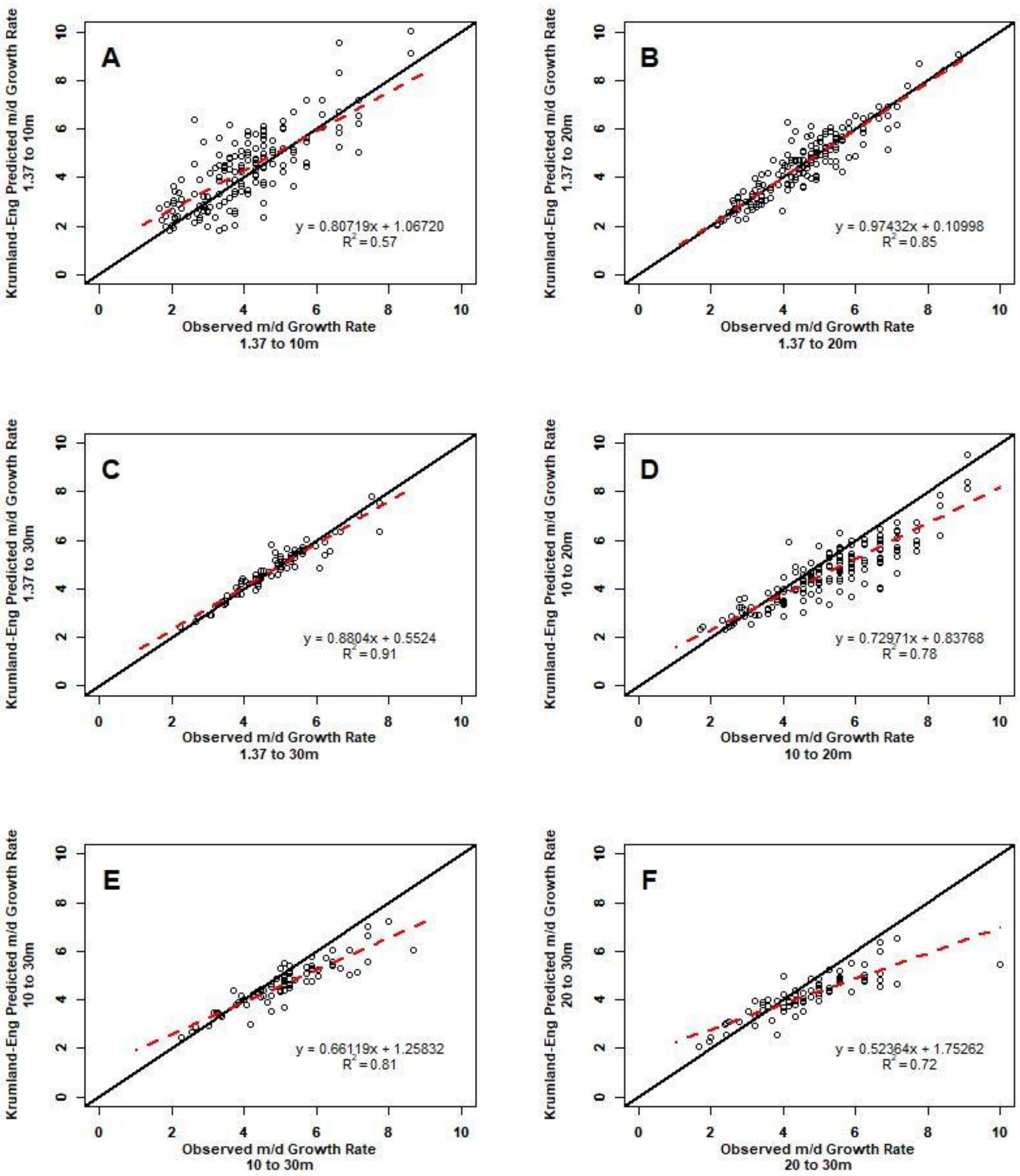

Figure 2.6. Relationship between Krumland and Eng (2005) predicted growth rates and observed growth rates.

Discussion

Proper landscape stratification via the use of an orthogonal sampling matrix provided an efficient way to ensure samples were collected across wide ranges of growth factors and site conditions in an unbiased manner (Forest Biometrics Research Institute, 2020; Hemingway and Kimsey 2020). Although site conditions varied greatly between sample locations, consistent patterns regarding accuracy of Krumland and Eng (2005) predicted growth rates for all stem segments across our study area allowed us to make a confident assessment of these site index approaches for this region.

The Krumland and Eng (2005) site index curve for interior Douglas-fir performed well when predicting BH to 20 and BH to 30 m growth rates. This is not overly surprising as tree length segments in essence reflects the one-point approach used in breast-height and total tree height models (i.e., Krumland and Eng, 2005). Thus, use of the DFI_CR2_MMC equation in this region remains informative, particularly when interested in growth rates relative to stands managed for longer-term ecosystem services or when user standard operating procedures incorporate growth and yield modeling applications that prefer one-point site index information.

For most tree segments, Krumland and Eng (2005) growth rate predictions overestimated growth of slower growing trees (<4 m/d), especially for BH to 10 m segments. Predicted values typically fell below observed values for middle and upper tree segments growing at 4 meters per decade or faster. This information and accurate tree-length (BH to 20 and BH to 30 m) growth rate predictions may serve as indication that compensation or front-loading of growth rates is likely taking place within the Krumland and Eng (2005) curves due to equation form. Alternatively, although we selected the most suitable site trees for our study, the quality of our site trees was reflective of the integrity of trees left on the landscape from previous management practices conducted by numerous landowners over time. Consequently, our sample trees may have differed in growth compared to those used to create the DFI_CR2_MMC equation.

Although predictions for BH to 20 and BH to 30 m segments were not significantly different than observed values, significant underpredictions were evident for 10 to 20, 10 to 30, and 20 to 30 m log segments. Limitations of one-point approaches are likely responsible for these underpredictions. One-point approaches, including Krumland and Eng (2005) site index, are built to quantify tree-length growth. The inherent nature of these approaches poses challenges when differentiating growth rates of smaller, position-specific stem segments as they cannot capture the nuance of site induced temporal effects on growth rates across these log segments.

On average, growth rates of 10 to 20 m segments, where trees reach merchantability, were underpredicted by 0.57 meters per decade. This underprediction can lead to significant misconceptions of a site's productive potential, particularly if site growth potential needs to be modeled by segment or over shorter time intervals. We observed in this study that trees are reaching desirable merchandizing heights in the second log significantly faster than predicted by Krumland and Eng (2005) curves, which in turn can lead to severe negative impacts on growth and yield, harvest scheduling, planting regimes, and timberland appraisals (Newell and Eves 2009; Latta et al. 2010; Devaranavadgi et al. 2013; Bontemps and Bouriaud 2014). We observed that only through a general overprediction of first log growth rates do Krumland and Eng (2005) full tree growth rates to 20 or 30 m achieve an overall similar growth rate to the two-point method. Consequently, tree segment growth differentials observed in this study between the one-point method (Krumland and Eng) and the two-point method coincide with other studies that support the two-point approach as a superior alternative to one-point approaches when assessing effects of site on log growth rates (Zeide 1978; Arney et al. 2009; Hemingway 2020).

Conclusion

Significant differences in segmented tree growth rates existed between one-point (Krumland and Eng (2005)) and two-point site index approaches. On average, Krumland and Eng (2005) site index overpredicted growth rates for breast-height to 10 m segments by 0.28 meters per decade and underpredicted for 10 to 20, 10 to 30, and 20 to 30 m segments by 0.57, 0.47, 0.46 meters per decade respectively. Predicted growth rates for breast-height to 20 and breast height to 30 m segments were consistent with observed growth rates, indicating that Krumland and Eng (2005) site index may remain informative when interested in tree-length growth rates or when considering long-term forest management practices consistent with ecosystem services other than intensive, short-term rotations. Because the two-point site index approach more accurately captured second log (10 to 20 m) growth rates, it may serve as a superior approach to quantify growth for intensively-managed stands with shorter rotation ages or where there is a need to more accurately define log segment growth rates as a function of site. Therefore, the overall applicability and effectiveness of both approaches for our study area is highly dependent on management practices and objectives.

Chapter 3: Modeling One and Multi-Point Douglas-fir Site Indices Using Machine Learning and Biological Growth Factors

Abstract

Incentive-driven demand for more accurate and reliable forest productivity estimates is endlessly increasing as carbon markets flourish and global need for forest products remains unabated. Forest productivity is widely used as a calibration parameter in growth and yield modeling applications, but not all applications are compatible with the same site indices. Additionally, methods previously used to spatially predict site indices, such as multi-linear regression (MLR) and geographically-weighted regression (GWR), are susceptible to problems regarding multicollinearity and small datasets. To address these issues, we explored the use of an extreme gradient boosting (XGBoost) machine learning algorithm to model one and multi-point site index approaches. Climatic, edaphic, and topographic predictors were used to predict Krumland and Eng (2005) site index (a one-point model) and 10-meter site index (10MSI) (a two-point model) across our study area. A landscape-wide stratification of known growth factors was essential to generating accurate and realistic estimates of productivity. Multiple statistical models were created and compared for each site index method. The lowest 10-fold cross-validated RMSE value was used to select final models for each method. Final models were used to generate 0.4 – hectare resolution raster layers of productivity for the entire study area.

Introduction

Forest productivity is an essential element in the operations and workflows of modern forest managers. Productivity estimates are highly impactful on all aspects of the forest industry, including growth and yield projections, real estate transactions, carbon markets, and property tax rates (Huston and Marland 2003; Seagle 2008; Newell and Eves 2009). The importance of productivity has created a steady increase in demand for evaluation of methods in which productivity estimates are derived, as well as improved levels of accuracy.

For decades, the estimated height of a forest stand at a given age, universally known as site index, has served as the dominant indicator of forest productive potential (Powers 1972; Hägglund and Lundmark 1977; Brown and Loewenstein 1978; Batho and Garcia 2006; Parresol et al. 2017). Amongst the most widely used site index methods are one-point approaches, which only require one height-age pair. Because one-point methods are heavily relied on by forest managers, they are widely used within growth and yield modeling applications, including many variants of the Forest Vegetation Simulation (FVS) growth and yield software platform (Batho and Garcia 2006b; Monserud et al. 2006; Kimsey et al. 2008; Hemingway 2020).

The most appropriate one-point inland Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco var. menziesii) site index equation for our study area was built by Krumland and Eng (2005) using stem analysis data. This equation is currently implemented operationally and is considered a reliable option when generating productivity estimates. Unfortunately, not all growth and yield models are compatible with one-point approaches. The 10m Site Index (10MSI) method is a two-point approach that is known to provide accurate site productivity estimates by eliminating biases from both early and late stage tree growth (Arney et al. 2009; Hemingway 2020). Two height-age measurements are used; one at 10 meters, and the other at 20 meters (Arney et al. 2009; Hemingway 2020). A growth rate expressed in meters per decade can be extracted from these measurements and used to define forest productivity (Arney et al. 2009; Hemingway 2020; Zeide, 1978), and is the backbone of growth and yield modeling in the widely used Forest Projection and Planning System (FPS) software package (Forest Biometrics Research Institute, 2020). Because standard operating procedures differ between landowners and managers in regard to growth and yield modeling, generating productivity estimates for both one and two-point site index approaches would likely prove beneficial.

Numerous studies suggest relationships between productivity and environmental factors may provide the most appropriate approach to obtain accurate predictions of forest productivity (Brown and Loewenstein 1978; Grier et al. 1989; Kimsey et al. 2008; Aertsen et al. 2012; Bontemps and Bouriaud 2014; Parresol et al. 2017; Hemingway 2020). Climate, edaphic properties, topography, and incoming solar radiation are amongst leading factors documented to influence tree growth (Corona et al. 1998; Skovsgaard and Vanclay 2008; DeYoung 2016). These growth factors and their interactions are likely responsible for a significant proportion of variance in productivity, making them suitable predictors when constructing models (Grier et al. 1989; DeYoung 2016).

Many modeling approaches, including multiple linear regression (MLR) and geographically-weighted regression (GWR), have been used to spatially estimate forest productivity with environmental growth factors over widespread landscapes (Monserud and Rehfeldt 1990; Kimsey et al. 2008). Although these parametric or semi-parametric approaches have proven themselves as generally effective, weaknesses arise when working with numerous autocorrelated predictor variables and small datasets (Woolery et al. 2002). Machine learning may provide a solution to these traditional problems.

Machine learning is a concept that has stepped to the center of attention for many statisticians and researchers. Gradient tree boosting (GTB) is perhaps one of the most powerful machine learning algorithms in use today, especially in terms of predictive capability (Nielsen 2016; Santhanam et al. 2016; Truong et al. 2020). Because GTB models can often be sufficiently trained with small datasets or with datasets containing missing values (Santhanam et al. 2016; Truong et al. 2020), they serve as a highly viable option when imposed with funding or time constraints. Additionally, the ensemble tree-based nature of GTB models result in reduced risk of overfitting and issues related to variable multicollinearity when used for prediction purposes (Dong et al. 2020).

Extreme Gradient Boosting (XGBoost) is a machine learning algorithm built on a boosted tree foundation that contains all the benefits of standard GTB models and more. Unlike the GTB algorithm, XGBoost takes a multi-threaded approach which optimizes the use of the machine's CPU core, resulting in improved efficiency, processing speed, and performance (Santhanam et al. 2016). Over the past several years, XGBoost models have declared superiority over other algorithms by winning numerous machine learning competitions (Nielsen 2016). Perhaps more importantly, XGBoost has become a powerful tool for real-world problem solving (Dong et al. 2020) and may serve as a reliable modeling approach for predicting forest productivity.

To meet the increased demand for accurate and readily available site productivity estimates to be used in various growth modeling applications, we sought to model both regional one and multi-point site index equations across our study area. We explored the use and performance of an XGBoost machine learning algorithm with climatic, edaphic, and topographic growth factors as predictors. Our research objectives were to 1) utilize machine learning and environmental growth factor predictors to construct Krumland and Eng (2005) site index and 10MSI Douglas-fir prediction models, 2) evaluate accuracy of constructed prediction models, and 3) use final prediction models to geospatially map Douglas-fir site indices across the study area of interest.

Methods and Materials

Study Area

A 66,283-hectare land ownership that covered parts of Shasta, Trinity, and Siskiyou counties of north central California, USA served as the study area (Fig. 3.1). Site conditions throughout the study area varied significantly. The range of elevations was 435 m to 2198 m above sea level with an average of 1237 m. The range of mean annual precipitation (MAP) was 586 mm to 1939 mm with an average of 1409 mm. Geologic soil parent materials ranged from ultramafic rocks and serpentine soils in the west, to volcanics located around Mt Shasta. The range of depths to restrictive layer was 0 cm to 251 cm with an average of 137 cm. Although Douglas-fir was a large species component, other species including white fir (*Abies concolor* (Gord.) Lemm var. lowiana), ponderosa pine (*Pinus ponderosa* Dougl. ex Laws. var. ponderosa), sugar pine (*Pinus lambertiana*), and California incense cedar (*Calocendrus decurrens*) were common within the study area.

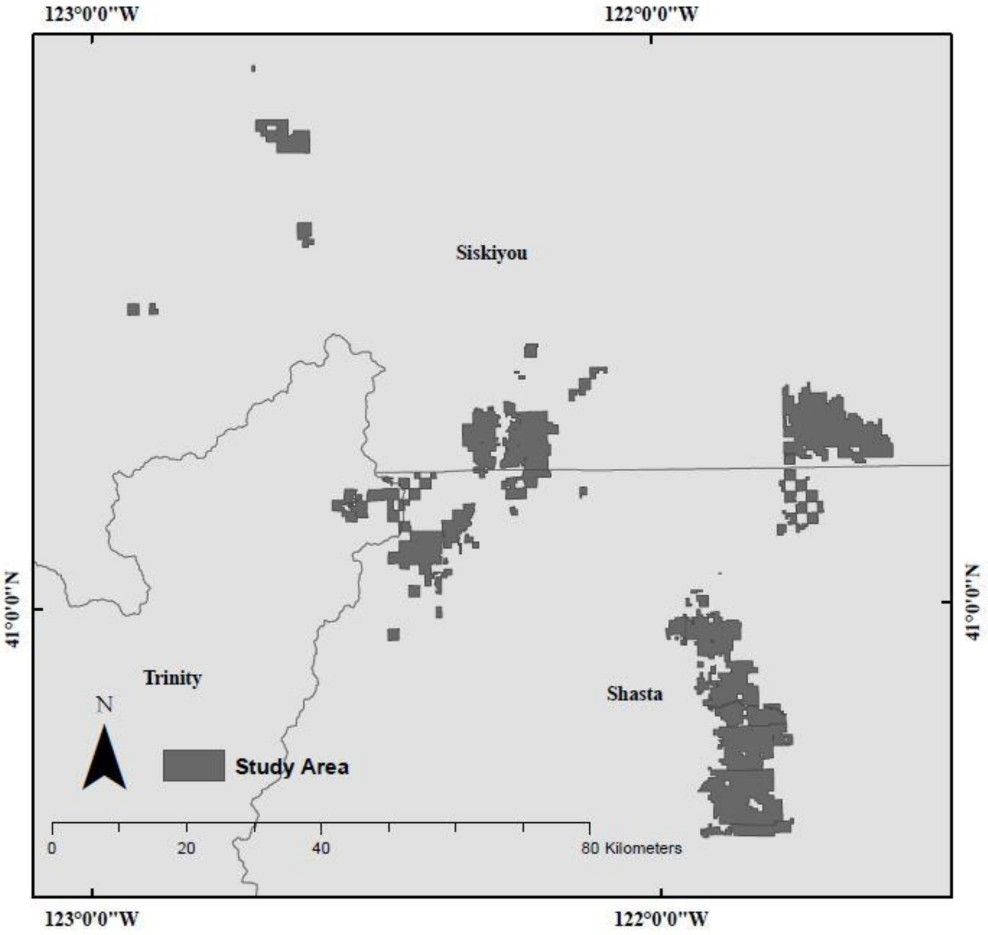

Figure 3.1. Study Area.

Stratification and Sample Selection

The study area was stratified by known growth factors related to temperature, moisture, and soil characteristics to ensure the vast range of present site conditions were represented by collected samples (Hemingway 2020). A 0.4-hectare point grid with attributes of mean annual precipitation (MAP), degree days above 5°C (DD5), solar radiation (RAD), depth to restrictive layer (DEP2RESLYR), available water supply from 0 to 100 cm (AWS100), geologic soil parent material type – a proxy for soil nutrient availability and development (SPM), and elevation (EL) was laid across the study area. The interaction of

DD5 and scaled RAD values served as surface Heat Load (HL) values. A single soil metric termed soil quality index (SQI) was generated using DEP2RESLYR, AWS100, and SPM values.

ClimateNA 1961-1990 30-yr normals were used to obtain MAP and DD5 values (Wang et al. 2016). ESRI's ArcMap 10.7.1 Points Solar Radiation tool was used to generate RAD values. Gridded National Soil Survey Geographic Database (gNATSGO) soil maps (Natural Resources Conservation Service, 2020) were used to obtain data for generating SQI values. A 30-m United States Geological Survey (USGS) digital elevation model (DEM) was used to obtain EL values. GIS extraction tools in ArcMap 10.7.1 were used to extract all soil and elevation data to gridded points. Point locations where Douglas-fir was not the dominant stand component, indicated by recent timber cruise data, were eliminated as potential sample locations.

27 unique strata were created with the use of an orthogonal sampling matrix and variables of MAP, HL, and SQI (Fig. 3.2; Arney et al. 2009; Hemingway and Kimsey 2020). Three strata bins of equal size for continuous variables of MAP and HL were defined by establishing breaks at the mean of each variable ± ½ standard deviation. A value of 1 was given to potential sampling points in the lowest bin (< ½ standard deviation), a value of 2 was given to points with values in the middle bin (± ½ standard deviation), and a value of 3 was given to points in the highest bin (> ½ standard deviation).

Douglas-fir site index values previously obtained from regional stand inventories were used as proxies for SPM quality (i.e., a means to develop a continuous proxy variable for SPM strata). Average site index values were calculated for each SPM and were then used to rank SPMs from least to most productive. The mean ± ½ standard deviation of averaged site index values was used to establish SPM strata breaks. A value of 1 was given to potential sampling points with SPMs in the lowest site index bin, a value of 2 was given to those with SPMs in the middle bin, and a value of 3 was given to those in the highest bin. Binning of DEP2RESLYR and AWS100 was conducted similar to MAP and HL by using the mean ± ½ standard deviation to establish breaks. DEP2RESLYR, AWS100, and SPM bin values were then summed and divided by three to obtain an average value for each point location.

Averaged values were then divided into three proportionate bins by establishing breaks at the mean ± ½ standard deviation. A SQI value of 1 was given to points that fell into the lowest bin, A SQI value of 2 was given to points that fell into the middle bin, and a SQI value of 3 was given to points that fell into the highest bin. Figure 3.3 displays the workflow for obtaining SQI values.

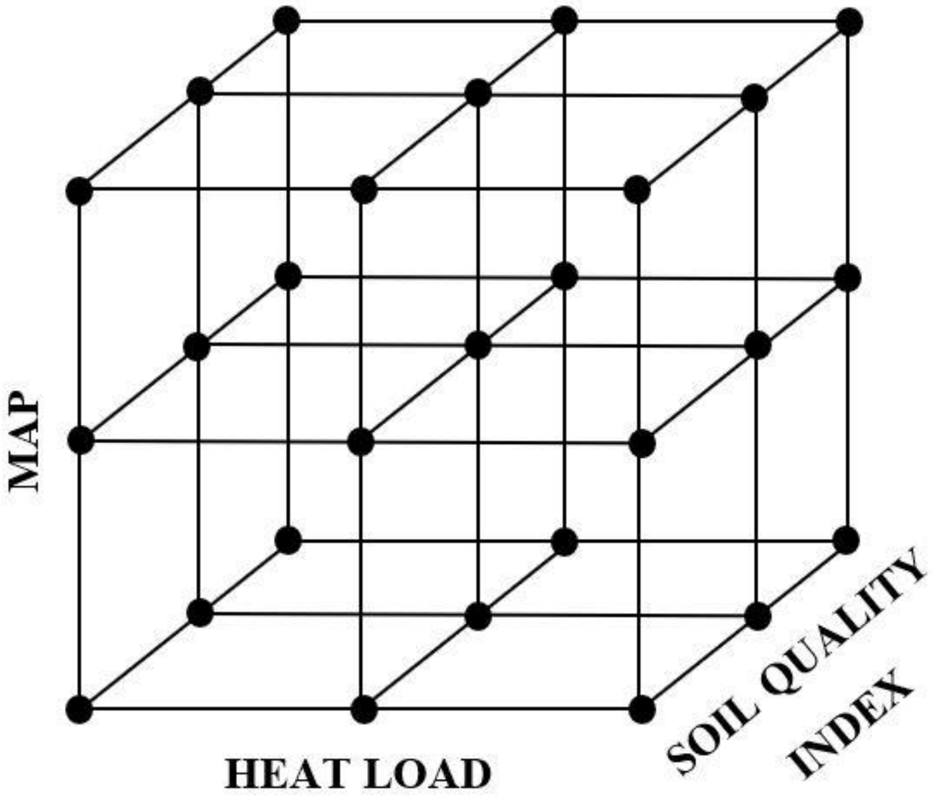

Figure 3.2. Orthogonal sampling matrix.

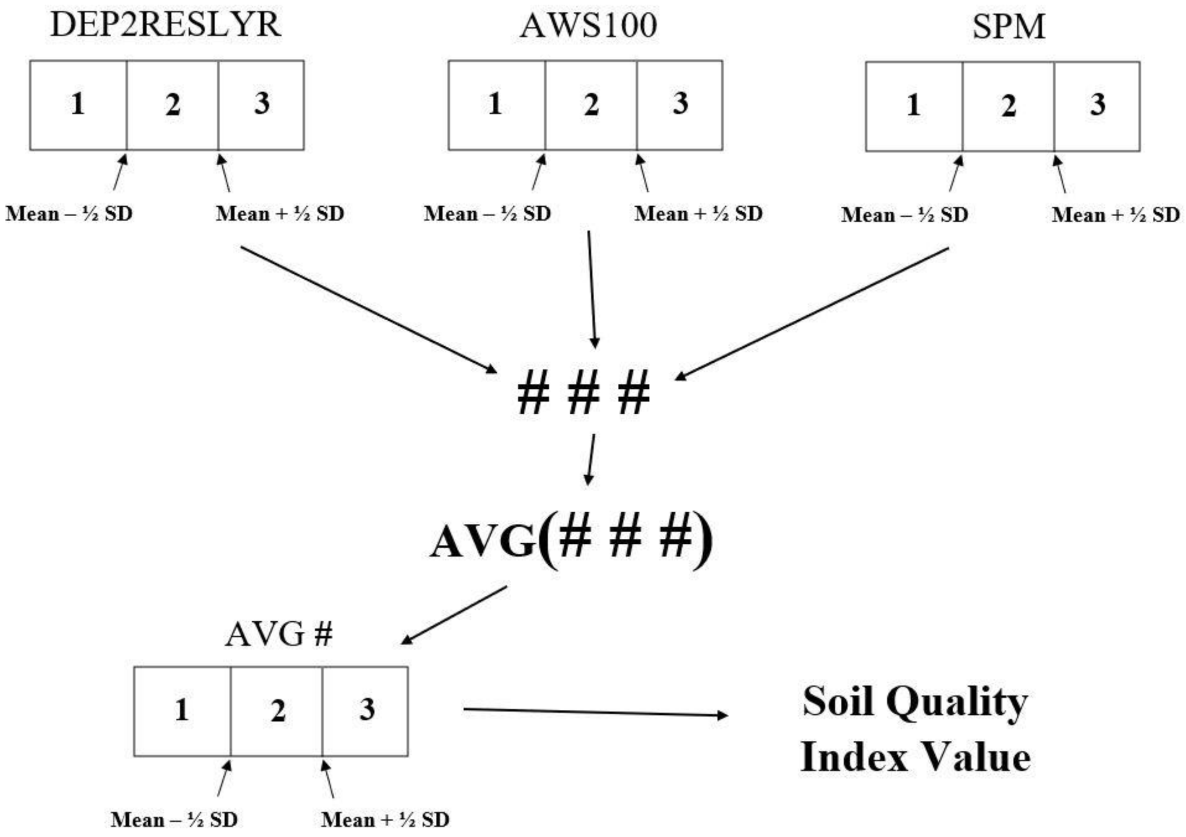

Figure 3.3. Workflow for obtaining soil quality index values.

The orthogonal sampling design using MAP, HL, and SQI was then replicated for low (<1200 m) and high (≥1200 m) elevations, resulting in a total of 54 strata. Stratifying in this manner allowed us to collect samples across the ranges of MAP, HL, SQI, and EL within the study area to unbiasedly represent variation in site conditions found on the landscape.

Sample locations were chosen at random across the 54 strata. Ocular inspections of selected sample locations were conducted during the summer of 2020. Inspections were important prior to sampling to verify if suitable Douglas-fir site trees existed at selected locations and whether or not such trees, if present, were accessible. To be considered a suitable site tree, a tree needed the following characteristics: 1) dominant or codominant position in the canopy, 2) greater than or equal to 20 m tall, 3) consistent growth from an early age – indicated by uniform spacing of branch whorls up the tree bole, 4) absent of defect or indication of illness, 5) growing in a non-advantageous location relative to surrounding trees, and 6) accessible for safe felling. If a sample location was determined to be insufficient, a new location for the same strata was chosen at random for inspection. In total, 81 locations representing diverse regional growing conditions were selected for sampling across the 54 strata (Fig. 3.4, Tables 3.1 and 3.2).

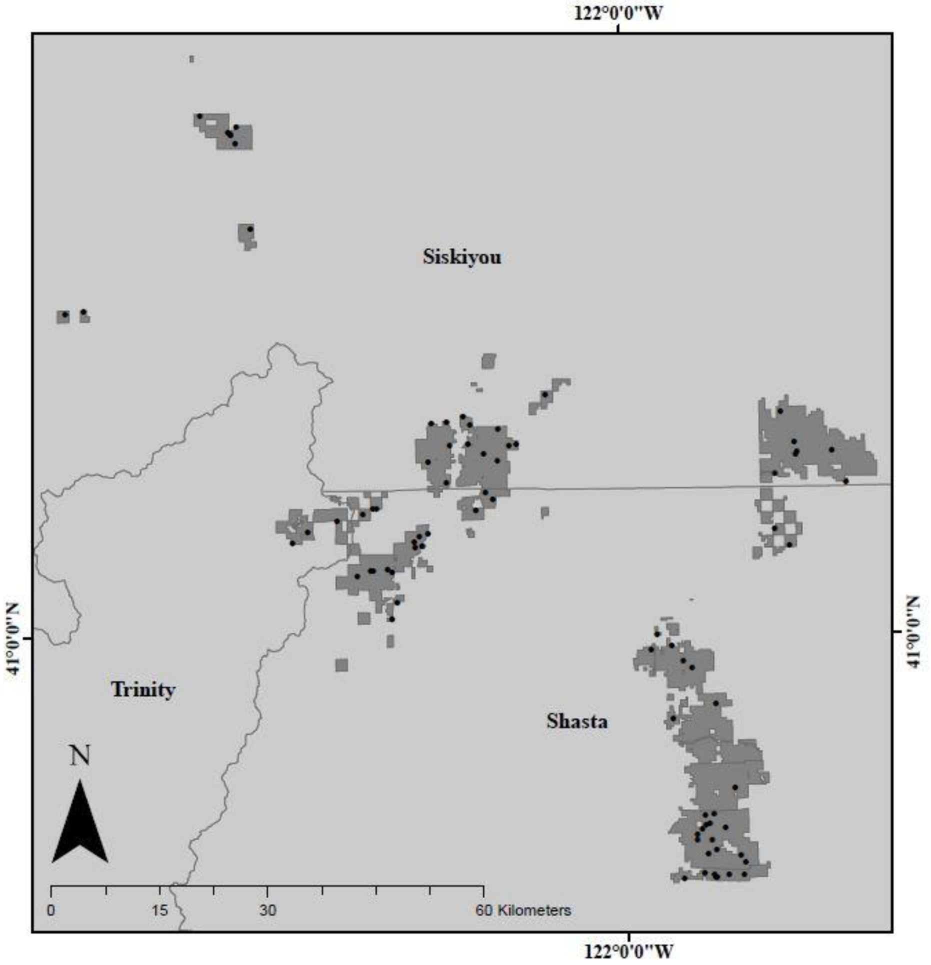

Figure 3.4. Spatial distribution of selected sample locations.

Table 3.1. Summary of site conditions for selected sample locations.

	Range	*Mean*	*SD*
EL[1] (m)	553 - 1596	1173	229
MAP[2] (mm)	637 - 1906	1396.4	359.9
DD5[3]	1878 - 3414	2583.2	356.6
RAD[4] (WH/m^2)	644035 - 1547840	1279950	209694
AWS100[5] (cm)	3.5 - 38.21	13.1	8.3
DEP2RESLYR[6] (cm)	23 - 201	129.8	50.8

[1] Elevation

[2] Mean annual precipitation

[3] Degree days greater than or equal to 5 °C

[4] Solar radiation

[5] Available water supply from ground surface to 100 cm

[6] Depth to restrictive layer

Table 3.2. Representation of geologic soil parent material types (SPM) amongst selected sample locations.

SPM	*% of Sample Locations*
Alluvium	1.2
Colluvium	2.5
Serpentine	4.9
Ash over volcanics	3.7
Volcanics	30.9
Tephra	11.1
Ultramafic	9.9
Igneous (unknown)	1.2
Igneous (extrusive)	3.7
Igneous (intrusive)	1.2
Metamorphic	28.4
Glacial	1.2

Sampling

Sampling of Douglas-fir trees was completed in the summer of 2021. In total, data was collected from 162 sample trees. Selected sample trees were felled and ages for cross sections at 10 and 20 m height positions were recorded. Ages were then used to calculate 10-meter site index (10MSI) using

$$\mathit{10MSI} = \frac{10_m * 10_{yrs}}{age_{10} - age_{20}} \qquad \text{[Eq. 3.1]}$$

where $\mathit{10MSI}$ is the observed meters per decade growth rate, 10_m is the difference in length between 10 and 20 m height positions, 10_{yrs} is the number of years in a decade, age_{10} is the age of the tree at 10 m, and age_{20} is the age of the tree at 20 m. If one of the two trees at each sample location did not have 10MSI value that fell within plus or minus 15 percent of the mean 10MSI value for the two trees, a third tree was selected and felled. Only data from the two most similar trees at each location were retained. 10MSI values for the two trees were then averaged to provide a 10MSI value for each sample location.

In addition to 10MSI, one-point site index values using the DFI_CR2_MMC (Krumland and Eng 2005) equation with a base age of 50 were calculated for each sample tree. The equation is written as

$$H = 4.5 + (H_0 - 4.5)\left\{\frac{[1 - exp(-0.01564 * A)]}{[1 - exp(-0.01564 * A_0)]}\right\}^{(-6.260 + 38.98/R_0)} \qquad \text{[Eq. 3.2]}$$

where H is site index, H_0 is observed total tree height, A is a specified base age, A_0 is observed age at BH (breast-height), and R_0 is an unobservable site productivity or growth intensity variable. R_0 can be calculated using

$$R_0 = \frac{\left(L_0 - (-6.260 * Y_0)\right) + \sqrt{\left(L_0 - (-6.260 * Y_0)\right)^2 - 4(38.98)Y_0}}{2} \qquad \text{[Eq. 3.3]}$$

letting

$$L_0 = ln(H_0 - 4.5) \qquad \text{[Eq. 3.4]}$$

$$Y_0 = ln\left(1 - exp(-0.01564 * A_0)\right) \qquad \text{[Eq. 3.5]}$$

where H_0 is observed total tree height and A_0 is observed age at BH. Similar to 10MSI values, Krumland and Eng (2005) site index values for each site were averaged to provide a sample location value.

Creating Krumland and Eng (2005) and 10MSI Site Index Prediction Models

In a similar manner, we created two extreme gradient boosting models (XGBoost) in R (R Core Team, 2021) to predict Krumland and Eng (2005) and 10MSI site index values. The first (KRUM/10MSI_1) used a random 80% of the sample data to train the model and the remaining 20% for validation. The second (KRUM/10MSI_2) used all sample data to train the model. Over 250 climatic, topographic, and edaphic predictor variables were used as inputs. Hyperparameters nrounds, eta, max_depth, colsample_bytree, subsample, gamma, and min_child_weight were adjusted via a tune grid to create optimal model fits. 10-fold cross validation error rate was used to select the final model. Probability density function (PDF) curves and absolute error (AE) quantiles were used to further assess model performance.

Generating Prediction Raster Layers

Krumland and Eng (2005) and 10MSI site index raster layers were generated for the entire study area using the best performing model for each method. The predict() function in R (R Core Team, 2021) was used to apply final models to a 0.4-hectare point grid with attributes of all model predictor variables. Point predictions were then rasterized in ArcMap 10.7.1 at 0.4-hectare resolution using the Point to Raster tool.

Results

Observed Krumland and Eng (2005) site index values varied from 11.14 to 39.35 m with a mean value of 22.94 m and a standard deviation of 5.82 m (Table 3.3). Observed 10MSI values for sample locations varied from 1.75 to 9.1 meters per decade with a mean value of 5.26 meters per decade and a standard deviation of 1.5 meters per decade (Table 3.3).

Table 3.3. Summary of observed Krumland and Eng (2005) site index and 10MSI values.

	Min	*Max*	*Mean*	*SD*	*Unit*
Krumland and Eng SI	11.14	39.35	22.94	5.82	m
10MSI	1.75	9.10	5.26	1.50	m/d

The KRUM_1 model yielded a 10-fold cross-validated RMSE of 5.86 m and a testing RMSE of 3.59 m on the withheld 20% of sample data (Table 3.4). The KRUM_2 model yielded a 10-fold cross-validated RMSE of 5.26 m (Table 3.4). Because all sample data were used to create the KRUM_2 model, an independent validation was not possible. Optimally-tuned hyperparameter values for Krumland and Eng (2005) site index models are shown in Table 3.5.

The 10MSI_1 model yielded a 10-fold cross-validated RMSE of 1.56 meters per decade and a testing RMSE of 1.18 meters per decade on the withheld 20% of sample data (Table 3.4). The 10MSI_2 model yielded a 10-fold cross-validated RMSE of 1.41 meters per decade (Table 3.4). Similar to KRUM_2, all sample data were used to create the 10MSI_2 model; therefore, an independent validation was not possible. Optimally tuned hyperparameter values for 10MSI models are shown in Table 3.5.

Table 3.4. RMSE values for prediction models.

	RMSE		
Model	*10-Fold CV*	*Ind. Validation*	*Unit*
KRUM_1	5.86	3.59	m
KRUM_2	5.26	NA	m
10MSI_1	1.57	1.18	m/d
10MSI_2	1.41	NA	m/d

Table 3.5. Tuned hyperparameter values for XGBoost prediction models.

	Model			
Hyperparameter	*KRUM_1*	*KRUM_2*	*10MSI_1*	*10MSI_2*
nrounds	500	500	200	500
eta	0.01	0.01	0.3	0.1
max_depth	10	5	10	5
colsample_bytree	0.5	0.5	0.7	0.7
subsample	0.5	0.5	1	0.7
gamma	10	0	10	10
min_child_weight	5	3	5	5

Although KRUM_1 and 10MSI_1 yielded low independent validation RMSE values, KRUM_2 and 10MSI_2 models were selected as final models based on 10-fold cross validated RMSE values. PDF curves showing distributions of KRUM_2 and 10MSI_2 predictions compared to observed values for the 81 sampled locations are shown in Figure 3.5.

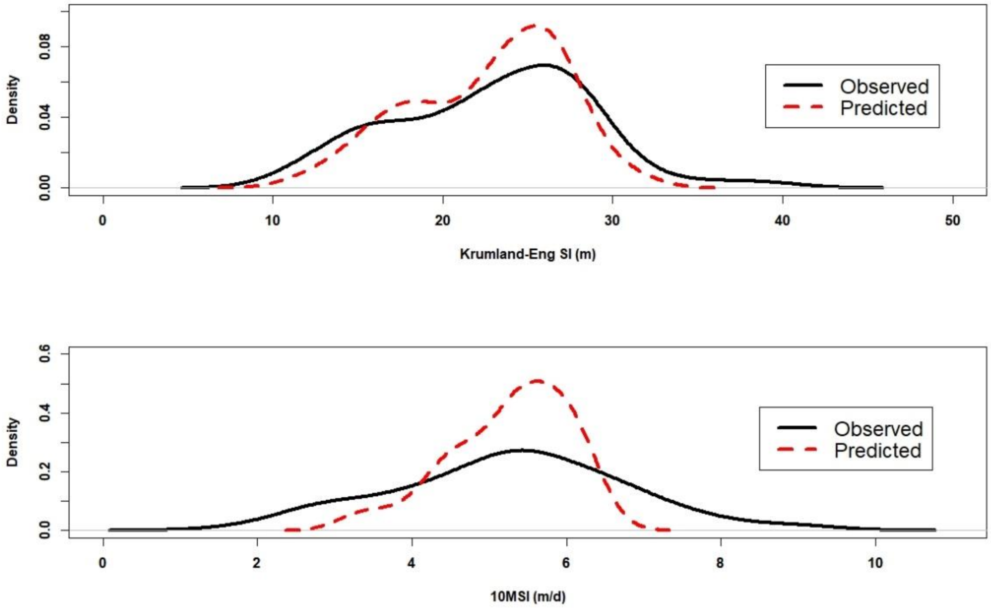

Figure 3.5. PDF curves for predicted and observed values for sampled locations by site index model.

Analysis of absolute error (AE) quantiles for the KRUM_2 model indicates that 50% or less of the observations were predicted to an accuracy of ± 0.87 m, 75% or less of the observations were predicted to an accuracy of ± 1.70 m, and 90% or less of the observations were predicted to an accuracy of ± 2.52 m (Table 3.6). Analysis of AE quantiles for the 10MSI_2 model indicates that 50% or less of the observations were predicted to an accuracy of ± 0.47 m/d, 75% or less of the observations were predicted to an accuracy of ± 1.17 m/d, and 90% or less of the observations were predicted to an accuracy of ± 1.61 m/d (Table 3.6).

Table 3.6. AE quantiles for final prediction models.

		AE Quantile		
Model	***Unit***	***50%***	***75%***	***90%***
KRUM_2	m	0.87	1.70	2.52
10MSI_2	m/d	0.47	1.17	1.61

When extrapolated to the entire study area, KRUM_2 predictions ranged from 12.6 to 30.5 m with a mean of 22.2 m and a standard deviation of 3.2 m. 10MSI_2 predictions for the entire study area ranged from 3.2 to 6.5 m/d with a mean of 5.3 m/d and a standard deviation of 0.7 m/d. Mapped Krumland and Eng (2005) and 10MSI site index predictions for the study area are shown in Figures 3.6 and 3.7.

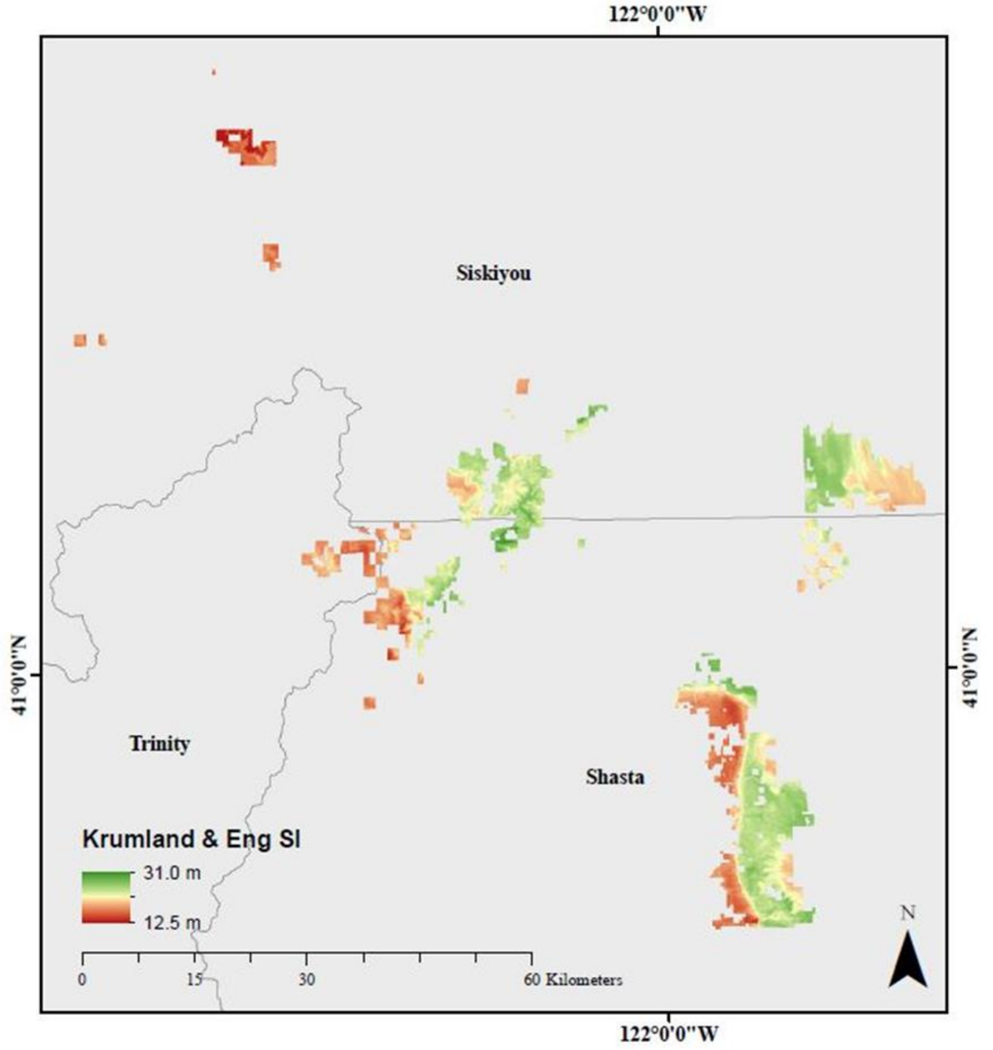

Figure 3.6. Map of Krumland and Eng (2005) site index predictions.

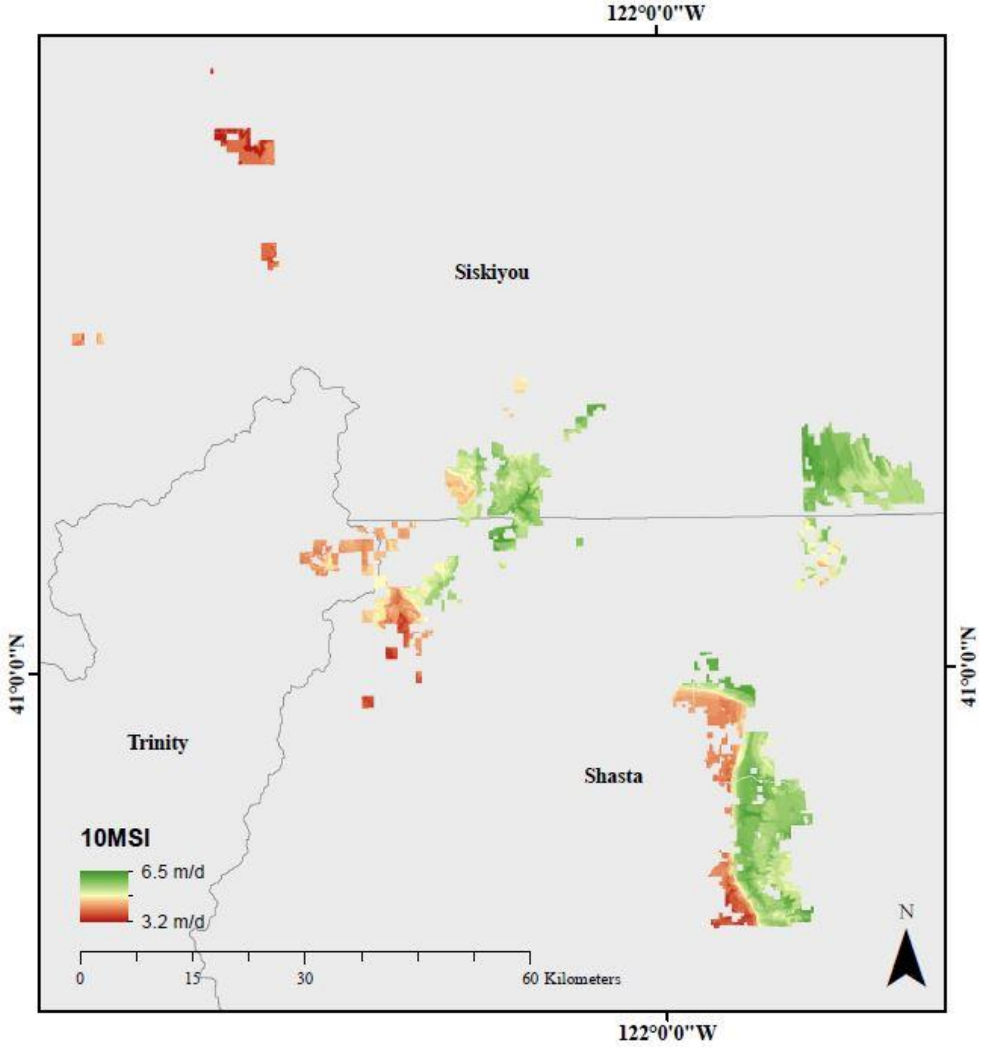

Figure 3.7. Map of 10MSI predictions.

Discussion

XGBoost prediction models constructed with climatic, edaphic, and topographic independent variables were successful in modeling Krumland and Eng (2005) and 10MSI site index across our study area with acceptable levels of error, often less than those reported using more traditional parametric or semi-parametric approaches (Kimsey et al, 2008). Proper landscape stratification via the use of an orthogonal sampling matrix was vital to ensure samples were collected across ranges of growth factors and site conditions in an unbiased manner as suggested by Hemingway (2020). This assured models were built to capture impacts on height growth resulting from various landscape conditions throughout our study area and provide the most accurate and realistic productivity estimates.

The implementation of a tune grid during model construction proved to be beneficial when determining optimal hyperparameter values. In addition to identifying the combination of hyperparameters that yielded the lowest 10-fold cross-validated RMSE, use of a tune grid provided the luxury of trying various combinations of hyperparameter values simultaneously. This resulted in significant time savings during the model construction process.

10-fold cross-validated models built with all observations were selected over models built with a random 80% of observations for several critical reasons. Even though independent test RMSE values for models built with 80% of the observations were acceptably low, those values can be misleading with relatively small datasets similar to our study. Independent test RMSE values are highly dependent on how data is randomly split into training (e.g., 80%) and testing (20%) portions, which needed to be considered in our “small” dataset of 81 observations. Random partitioning of the data can significantly change optimal hyperparameter values when working with small datasets. Inevitably, this leads to fluctuation in error values. This becomes especially problematic when extreme or unique values exist in the dataset. Issues arise when a model attempts to make a prediction on an extreme value using non-extreme training data, or when making a prediction on a non-extreme value using extreme training data. 10-fold cross-validated models built with all observations mitigated risks associated with extreme or unique observations because error statistics were averaged across 10 testing folds. Additionally, this allowed us to utilize all observations in model training, which we found beneficial when working with smaller datasets.

PDF curves shown in Figure 3.5 indicate that final models, especially KRUM_2, are performing acceptably statistically for a majority of locations. Table 3.5 indicates up to 75% of locations predicted by KRUM_2 and 10MSI_2 yield productivity estimates with residuals of less than ±1.7 m and 1.17 m/d respectively. In terms of the study area, that translates to approximately 49,712 hectares (or 75% of the land base) with productivity estimates to those levels of accuracy. Model performance was reduced for sites with very low or very high productivity, especially for the 10MSI_2 model. This could be the result of low sampling intensity in very low and very high productivity regions of the study area (i.e., an artifact of using ± ½ standard deviation for strata bins) and/or overlapping site variables across a range of site productivities. Because very few sites used to build models expressed extreme productivity (low or high) in the second log (10 to 20 m segment), the models likely had difficulty making confident predictions for such locations. Therefore, models default to predicted values that are supported by a majority of the dataset. This is supported by higher 90% AE quantile values for both models in Table 3.5. Reduced performance of 10MSI_2 when compared to KRUM_2 in regard to extreme sites may suggest additional sources of variation impacting second log growth exist beyond the specific predictors used in this study. Additionally, the use of ± 1 standard deviation for defining strata bins may have better captured these extreme values and improved predictive performance.

Accurate predictions for a majority of our study area support other studies that suggest the use of climatic, edaphic, and topographic factors to estimate forest productivity across widespread geographic regions (Brown and Loewenstein 1978; Grier et al. 1989; Kimsey et al. 2008; Aertsen et al. 2012; Bontemps and Bouriaud 2014; Parresol et al. 2017; Hemingway 2020). In total, 251 site variables were used as predictors when constructing models in attempt to explain as much variance as possible in Krumland and Eng (2005) and 10MSI site index. For many traditional modeling approaches that prefer reduced variable datasets, this would be overwhelming and likely penalize results. However, the nature of XGBoost allowed us to use more than 250 predictors, many of which were autocorrelated, with minimal associated risks (Dong et al. 2020). Although interpretability of specific cause and effect relationships may have decreased, predictive performance remained unaffected. Collectively, cross validation, the use of a tune grid, and the mechanisms behind the XGBoost algorithm mitigated risks of overfitting models.

Because productivity estimates are required for nearly all growth and yield modeling applications like FVS and FPS, but not all applications prefer the same productivity indices, it was important to build models and generate predictions for both Krumland and Eng (2005) site index and 10MSI. This allows local users to select between one and two-point site index methods according to their preference in growth and yield modeling software. Having predictions for both site index approaches may also prove advantageous if users find that one approach more accurately reflects productivity in certain management scenarios when compared to the other approach.

Conclusion

Correctly tuned XGBoost models using climatic, edaphic, and topographic predictors proved to be a successful approach to accurately predict site productivity across our study area. Up to 75% of the study area was predicted to ±1.7 m and ±1.17 m/d for Krumland and Eng (2005) site index and 10MSI respectively. Proper landscape stratification and sample selection were imperative to ensure samples were collected in an unbiased manner across a wide range of growth factors to yield the most accurate and realistic estimates. Implementation of a tune grid provided significant time savings when optimizing model hyperparameters. Including all observations in model training reduced 10-fold cross-validated RMSE values. Although, models built with 80% of observations yielded low independent testing RMSE values, error statistics were highly dependent on how the data was randomly partitioned into training and testing datasets.

Because regional growth and yield modeling applications prefer one-point site index and others prefer two-point site index, it was advantageous to generate predictions for both Krumland and Eng (2005) site index and 10MSI. Rasterized predictions at 0.4-hectare resolution were beneficial when visualizing changes in productivity across the study area, and they serve as informative data layers for local forest managers. Potential improvements to models and estimates are possible in the future with collection of additional Krumland and Eng (2005) site index and 10MSI data.

Chapter 4: Conclusion

Significant differences existed in segmented tree growth rates between one and two-point site index approaches. Krumland and Eng (2005) site index overpredicted growth rates for breast-height to 10 m segments and significantly underpredicted growth rates for 10 to 20, 10 to 30, and 20 to 30 m segments. However, significant differences in growth rates between approaches did not exist for breast-height to 20 and breast height to 30 m segments. This may suggest Krumland and Eng (2005) site index remains sufficient when interested in tree-length or long-term growth rates. The two-point site index approach more accurately captured second log (10 to 20 m) growth rates, suggesting it may serve as a superior approach when quantifying growth for intensively-managed stands with shorter rotation ages. Therefore, it can be concluded that overall applicability and effectiveness of both approaches for our study area is highly dependent on management practices and objectives.

XGBoost models using climatic, edaphic, and topographic predictors proved to be successful in accurately predicting site productivity across our study area. A widespread landscape stratification of known growth factors was essential to the study design and ensured models created statistically valid and accurate predictions. The use of a tune grid proved to be effective and time efficient when optimizing model hyperparameters. Models trained with all observations (KRUM_2 and 10MSI_2) yielded reduced 10-fold cross-validated RMSE values when compared to models built with 80% of observations (KRUM_1 and 10MSI_1). Although KRUM_1 and 10MSI_1 yielded attractive independent testing RMSE values, error statistics were highly dependent on the random partitioning of observations.

Predictive performance of both final models, assessed by PDF curves and AE quantiles, was sufficiently accurate for a majority of sampling locations. However, reduced accuracies were prevalent for sites with extremely low or high productivity. This may be the effect of low sampling intensity for locations of this nature (i.e., a product of using ± ½ standard deviation for defining strata bins). Additionally, this could be the effect of shared site conditions across a range of observed site productivities. Because the presence of extreme productivity values (low or high) was very minimal in training datasets, models likely struggled to make confident predictions for such locations. Consequently, the models defaulted to predictions that were supported by a majority of the dataset.

It was advantageous to generate both one-point (Krumland and Eng (2005) site index) and two-point (10MSI) prediction models for our study area because many growth and yield modeling applications that use site index data as a calibration parameter, such as FVS and FPS, have preferences to either one or two-point values. Rasterized predictions made with final models at 0.4-hectare resolution were beneficial when visualizing changes in productivity across the study area, and they serve as informative data layers for local forest managers. Potential improvements to prediction models and estimates are possible in the future with collection of additional Krumland and Eng (2005) site index and 10MSI data.

Literature Cited

Aertsen, W., V. Kint, K. Von Wilpert, D. Zirlewagen, B. Muys, and J. Van Orshoven. 2012. Comparison of location-based, attribute-based and hybrid regionalization techniques for mapping forest site productivity. Forestry. 85(4):539–550.

Arney, J. D., B. L. Kleinhenz, and K. S. Milner. 2009. Estimating forest productivity: the 10m Site Index Method. Portland, OR.

Baker, F. S. 1944. Mountain climates of the western United States. Ecol. Monogr. 14(2):233.

Batho, A., and O. Garcia. 2006. De Perthius and the origins of site index: a historical note. FBMIS. 1:1–10.

Bontemps, J. D., and O. Bouriaud. 2014. Predictive approaches to forest site productivity: recent trends, challenges and future perspectives. Forestry. 87(1):109–128.

Brown, H. G., and H. Loewenstein. 1978. Predicting site productivity of mixed conifer stands in northern Idaho from soil and topographic variables. Soil Sci. Soc. Am. J. 42:967–971.

Bruce, D. 1926. A method of preparing timber-yield tables. J. Agric. Res. 32(6):543–557.

Carmean, W. H. 1954. Site quality for Douglas-fir in southwestern Washington and its relationship to precipitation, elevation, and physical soil properties. Soil Sci. Soc. Am. J. 18(3):330–334.

Cochran, P. H. 1979. Site index and height growth curves for managed, even-aged stands of Douglas-fir east of the Cascades in Oregon and Washington. Portland, OR.

Corona, P., R. Scotti, and N. Tarchiani. 1998. Relationship between environmental factors and site index in Douglas-fir plantations in central Italy. For. Ecol. Manage. 110(1–3):195–207.

Curtis, R. O. 1964. A stem-analysis approach to site-index curves. For. Sci. 10(2):241–256.

Devaranavadgi, S. B., S. Bassappa, and S. Y. Wali. 2013. Height-age growth curve modelling for different tree species in drylands of north Karnataka. Glob. J. Sci. Front. Res. Agric. Vet. Sci. 13(1).

DeYoung, J. 2016. Forest measurements: an applied approach. Open Oregon Educational Resources. 161–164 p.

Dolph, K. L. 1988. Site index curves for young-growth California White fir on the western slopes of the Sierra Nevada. Berkeley, CA.

Dong, W., Y. Huang, B. Lehane, and G. Ma. 2020. XGBoost algorithm-based prediction of concrete electrical resistivity for structural health monitoring. Autom. Constr. 114.

Dunning, D., and L. H. Reineke. 1933. Preliminary yield tables for second-growth stands in the California pine region. Washington, D. C.

Fernholz, K. 2007. TIMOs & REITs: what, why, & how they might impact sustainable forestry. Minneapolis, MN.

Forest Biometrics Research Institute, 2020. Forest Projection and Planning Software.

Grier, C. C., K. M. Lee, N. M. Nadkarni, G. O. Klock, and P. J. Edgerton. 1989. Productivity of forests of the United States and its relation to soil and site factors and management practices: a review. Portland, OR.

Hägglund, B., and J.-E. Lundmark. 1977. Site index estimation by means of site properties of Scots pine and Norway spruce in Sweden.

Hann, D. W., and J. A. Scrivani. 1987. Dominant-height-growth and site-index equations for Douglas-fir and Ponderosa pine in southwest Oregon. Corvallis, OR.

Hemingway, H. J. 2020. Defining and estimating forest productivity using multi-point measures and a nonparametric approach. University of Idaho.

Hemingway, H., and M. Kimsey. 2020. Estimating forest productivity using site characteristics, multipoint measures, and a nonparametric approach. For. Sci. 66(6):645–652.

Huston, M. A., and G. Marland. 2003. Carbon management and biodiversity. J. Environ. Manage. 67(1):77–86.

Kimsey, M. J., J. Moore, and P. McDaniel. 2008. A geographically weighted regression analysis of Douglas-fir site index in north central Idaho. For. Sci. 54(3):356–366.

Klemperer, W. D. 1976. Impacts of tax alternatives on forest values and investment. Land Econ. 52(2):135–157.

Krumland, B., and H. Eng. 2005. Site index systems for mahor young-growth forest and woodland species in northern California. Calif. For. 4:1–220.

Latta, G., H. Temesgen, D. Adams, and T. Barrett. 2010. Forest ecology and management analysis of potential impacts of climate change on forests of the United States pacific northwest. For. Ecol. Manage. 259:720–729.

Monserud, R. A. 1984. Height growth and site index curves for inland Douglas-fir based on stem analysis data and forest habitat type. For. Sci. 30(4):943–965.

Monserud, R. A. ., and G. E. Rehfeldt. 1990. Genetic and environmental components of variation of site index in inland Douglas-fir. For. Sci. 36(1):1–9.

Monserud, R. A., S. Huang, and Y. Yang. 2006. Predicting lodgepole pine site index from climatic parameters in Alberta. For. Chron. 82(4):562–571.

Newell, G., and C. Eves. 2009. The role of U. S. timberland in real estate. J. Real Estate Portf. Manag. 15(1):95–106.

Nielsen, D. 2016. Tree boosting with XGBoost. Norwegian University of Science and Technology.

Parresol, B. R., D. A. Scott, S. J. Zarnoch, L. A. Edwards, and J. I. Blake. 2017. Modeling forest site productivity using mapped geospatial attributes within a South Carolina Landscape, USA. For. Ecol. Manage. 406(October):196–207.

Powers, R. F. 1972. Estimating site index of Ponderosa pine in northern California...standard curves, soil series, stem analysis. Berkeley, CA.

R Core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Santhanam, R., N. Uzir, S. Raman, and S. Banerjee. 2016. Experimenting XGBoost algorithm for prediction and classification of different datasets. Int. J. Control Theory Appl. 9(40):651–662.

Seagle, S. W. 2008. Spatial patterns of potential forest productivity in Maryland: exploring landscape planning for forest carbon sequestraton and fragmentation reduction. Appalachian State University.

Skovsgaard, J. P., and J. K. Vanclay. 2008. Forest site productivity: A review of the evolution of dendrometric concepts for even-aged stands. Forestry. 81(1):13–31.

Soil Survey Staff. Gridded National Soil Survey Geographic (gNATSGO) Database for California. United States Department of Agriculture, Natural Resources Conservation Service. https://nrcs.app.box.com/v/soils. December 1, 2020 *(*FY2020 official release).

Truong, V. H., Q. V. Vu, H. T. Thai, and M. H. Ha. 2020. A robust method for safety evaluation of steel trusses using Gradient Tree Boosting algorithm. Adv. Eng. Softw. 147.

Vanclay, J. K. 1994. Modelling forest growth and yield: applications to mixed tropical forests. CAB International, Wallingford UK. 1–13 p.

Wang, T., Hamann, A., Spittlehouse, D., Carroll, C. 2016. Locally downscaled and spatially customizable climate data for historical and future periods for North America. PLoS ONE 11(6): e0156720. doi:10.1371/journal.pone.0156720

Weiner, J., and S. C. Thomas. 2001. The nature of tree growth and the age-related decline in forest productivity. Oikos. 94(2):374–376.

Weiskittel, A. R., N. L. Crookston, and P. J. Radtke. 2011. Linking climate, gross primary productivity, and site index across forests of the western United States. Can. J. For. Res. 41:1710–1721.

Woolery, M. E., K. R. Olson, J. O. Dawson, and G. Bollero. 2002. Using soil properties to predict forest productivity in Southern Illinois. J. Soil Water Conserv. 57(1):37–45.

Zeide, B. 1978. Standardization of growth curves. J. For. :289–292.