

Leveraging Open Data to Support
Forest Mapping, Modeling, and Policy Analysis
in the Pacific Northwest, USA

David Daniel Diaz

A dissertation

submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Gregory J. Ettl, Chair
Beth Gardner
Andrew N. Gray

Program Authorized to Offer Degree:

Environmental and Forest Sciences

© Copyright 2023

David Daniel Diaz

University of Washington

Abstract

**Leveraging Open Data to Support
Forest Mapping, Modeling, and Policy Analysis
across the Pacific Northwest, USA**

David Daniel Diaz

Chair of the Supervisory Committee:
Associate Professor, Gregory J. Ettl
School of Environmental and Forest Sciences

Increasing access to open data can offer unprecedented opportunities to conduct and share actionable science involving forest ecology, management, and conservation. This dissertation recounts three diverse applications of open data and ecosystem data science in the Pacific Northwest, USA: (i) to generate maps of forest conditions that can be easily updated as new earth observation data arrive; (ii) to develop a reproducible workflow to recalibrate the Forest Vegetation Simulator (FVS), a widely used growth-and-yield model, as new field observations arise; and (iii) to quantify tradeoffs in timber, carbon, and cash flow among alternative management approaches in Douglas-fir forests. Compiling a multi-agency dataset across Oregon and Washington, I demonstrate the potential for modern Machine Learning algorithms to effectively learn and generalize relationships between forest structure attributes with remote sensing across multiple ecoregions. Utilizing an extensive network of repeated forest inventory measurements, I highlight opportunities to simplify FVS equations and software maintenance while improving predictions of diameter growth for sixteen diverse tree species and learning

growth drivers simultaneously across multiple ecoregions from coastal Washington through California. Finally, by translating an open dataset depicting forest conditions across the landscape to stand-level units, I employ more than 250,000 simulations of forest management across western Oregon and Washington to conduct a tradeoff analysis involving varying rotation lengths, levels of green tree retention, and extent of riparian buffers. I find that forest management following Forest Stewardship Council requirements is likely to produce substantially higher carbon storage in forests and in wood products at a significant opportunity cost in terms of net present value of timber revenue compared to business-as-usual management following state forest practice regulations in both Oregon and Washington. Together, these three applications illustrate common use cases for forest data and technology to enhance decision-making capacity and understanding of current forest conditions, how confidently we can model forest dynamics, and how alternative management approaches can lead to important tradeoffs relevant to forest owners, managers, policymakers, and consumers of forest products.

TABLE OF CONTENTS

Chapter 1. Tradeoffs in Timber, Carbon, and Cash Flow under Alternative Management Systems for Douglas-Fir in the Pacific Northwest.....	1
1.0 Abstract.....	1
1.1 Introduction.....	2
1.2 Materials and Methods.....	7
1.3 Results.....	21
1.4 Discussion	28
1.5 Conclusions.....	33
1.6 Supplementary Materials	35
Chapter 2. Predicting forest structure attributes from lidar or satellite data and assessing model transferability across multiple ecoregions	53
2.1 Introduction.....	53
2.2 Methods.....	61
2.3 Results.....	76
2.4 Discussion	106
2.5 Future study	109
2.6 Conclusion	111
2.7 Supplementary Materials	111

Chapter 3. A reproducible framework to refit Forest Vegetation Simulator growth equations simultaneously across multiple ecoregions.....	133
3.1 Introduction.....	133
3.2 Methods.....	143
3.3 Results.....	161
3.4 Discussion	176
3.5 Future study	179
3.6 Conclusion	180
3.7 Supplementary Materials	181
REFERENCES CITED	212

LIST OF FIGURES

Figure 1.1. Study area.	3
Figure 1.2. Division of landscape into management units to be simulated.	8
Figure 1.3. Correspondence between selected parcels and the surrounding landscape....	10
Figure 1.4. Area encumbered by Riparian Management Zones under FPA and FSC rules for Oregon and Washington.	14
Figure 1.5. Carbon stored in three primary categories of pools/emissions under four alternative management scenarios.	22
Figure 1.6. Dynamics of three KPIs over 100 years under four alternative management scenarios.	23
Figure 1.7. Shifts in timber yield, carbon storage, and cash flow relative to business-as-usual (NPV~FPA) for each of the three alternative scenarios.	27
Figure S1.8. Default FVS Parameterization comparing PN variant output to measured stands (Part 1 of 3).	38
Figure S1.9. Default FVS Parameterization (Part 2 of 3).	39
Figure S1.10. Default FVS Parameterization (Part 3 of 3).	40
Figure S1.11. Modified FVS Parameterization (Part 1 of 3).	41
Figure S1.12. Modified FVS Parameterization (Part 2 of 3).	42
Figure S1.13. Modified FVS Parameterization (Part 3 of 3).	43
Figure S1.14. Comparisons of the area encumbered by FSC and FPA RMZs for Oregon and Washington.	49
Figure 2.1. Locations of remote sensing forest plots.	62
Figure 2.2. Distributions of forest structure variables across ecoregions comparing training datasets with all forested areas in Oregon and Washington.....	64
Figure 2.3. Illustration of sample re-weighting for domain adaptation	71
Figure 2.4. Matching FIA distributions with sample re-weighting	71
Figure 2.5. Illustration of nested cross validation for model tuning and validation.	73

Figure 2.6. Rankings of forest structure predictive models based on NRMSE (part 1 of 2).	78
Figure 2.7. Rankings of forest structure predictive models based on NRMSE (part 2 of 2).	79
Figure 2.8. Predictive performance (Normalized Root Mean Square Error, NRMSE) of lidar- and satellite-based models on three forest structure attributes (1 of 2).	80
Figure 2.9. Predictive performance (Normalized Root Mean Square Error, NRMSE) of lidar- and satellite-based models on three forest structure attributes (2 of 2).	81
Figure 2.10. Performance (Normalized Root Mean Square Error, NRMSE) of GBM-GLOBAL and GBM-INSIDER lidar models across all ecoregions.	82
Figure 2.11. Performance (Normalized Root Mean Square Error, NRMSE) of GBM-GLOBAL and GBM-INSIDER satellite models across all ecoregions.	83
Figure 2.12. Rankings of forest structure models based on relative bias (part 1 of 2).	85
Figure 2.13. Rankings of forest structure models based on relative bias (part 2 of 2).	86
Figure 2.14. Predictive performance (% bias) of lidar- and satellite-based models on three forest structure attributes (1 of 2).	87
Figure 2.15. Predictive performance (% bias) of lidar- and satellite-based models on three forest structure attributes (2 of 2).	88
Figure 2.16. Performance (% bias) of GBM-GLOBAL and GBM-INSIDER lidar models across all ecoregions.	89
Figure 2.17. Performance (% bias) of GBM-GLOBAL and GBM-INSIDER satellite models across all ecoregions.	90
Figure 2.18. Correspondence between predictions and observations for top-performing satellite and lidar models (1 of 2).	92
Figure 2.19. Correspondence between predictions and observations for top-performing satellite and lidar models (2 of 2).	93
Figure 2.20. Correspondence between predictions and observations for linear satellite and lidar models (1 of 2).	94
Figure 2.21. Correspondence between predictions and observations for linear satellite and lidar models (2 of 2).	95
Figure 2.22. Visitor transferability for canopy cover predictions.	99
Figure 2.23. Visitor transferability for aboveground live carbon predictions.	100

Figure 2.24. Effect of sample weighting on Visitor transferability for canopy cover predictions.	102
Figure 2.25. Effect of sample weighting on Visitor transferability for aboveground live carbon predictions.	103
Figure 2.26. Effect of sample weighting on Visitor transferability for canopy cover predictions.	104
Figure 2.27. Effect of sample weighting on Visitor transferability for aboveground live carbon predictions.	105
Figure S2.28. Improved predictive performance (Normalized Root Mean Square Error, NRMSE) with increasing training data for three structure attributes (1 of 2).	112
Figure S2.29. Improved predictive performance (Normalized Root Mean Square Error, NRMSE) with increasing training data for three structure attributes (2 of 2).	113
Figure S2.30. Rankings of forest structure predictive models based on mean average percent error.	114
Figure S2.31. Predictive performance (Mean Absolute Percent Error, MAPE) of lidar- and satellite-based models on three forest structure attributes (1 of 2).	115
Figure S2.32. Predictive performance (Mean Absolute Percent Error, MAPE) of lidar- and satellite-based models on three forest structure attributes (2 of 2).	116
Figure S2.33. Performance (Normalized Root Mean Square Error, NRMSE) of ElasticNet- GLOBAL and ElasticNet-INSIDER lidar models across all ecoregions.	117
Figure S2.34. Performance (Normalized Root Mean Square Error, NRMSE) of ElasticNet- GLOBAL and ElasticNet-INSIDER satellite models across all ecoregions.	118
Figure S2.35. Performance (Normalized Root Mean Square Error, NRMSE) of LASSO- GLOBAL and LASSO-INSIDER lidar models across all ecoregions.	119
Figure S2.36. Performance (Normalized Root Mean Square Error, NRMSE) of LASSO- GLOBAL and LASSO-INSIDER satellite models across all ecoregions.	120
Figure S2.37. Performance (Normalized Root Mean Square Error, NRMSE) of KNN-GLOBAL and KNN-INSIDER lidar models across all ecoregions.	121
Figure S2.38. Performance (Normalized Root Mean Square Error, NRMSE) of KNN-GLOBAL and KNN-INSIDER satellite models across all ecoregions.	122

Figure S2.39. Performance (Normalized Root Mean Square Error, NRMSE) of RF-GLOBAL and RF-INSIDER lidar models across all ecoregions.....	123
Figure S2.40. Performance (Normalized Root Mean Square Error, NRMSE) of RF-GLOBAL and RF-INSIDER satellite models across all ecoregions.....	124
Figure S2.41. Performance (% bias) of ElasticNet-GLOBAL and ElasticNet-INSIDER lidar models across all ecoregions.....	125
Figure S2.42. Performance (% bias) of ElasticNet-GLOBAL and ElasticNet-INSIDER satellite models across all ecoregions.....	126
Figure S2.43. Performance (% bias) of LASSO-GLOBAL and LASSO-INSIDER lidar models across all ecoregions.	127
Figure S2.44. Performance (% bias) of LASSO-GLOBAL and LASSO-INSIDER satellite models across all ecoregions.....	128
Figure S2.45. Performance (% bias) of KNN-GLOBAL and KNN-INSIDER lidar models across all ecoregions.	129
Figure S2.46. Performance (% bias) of KNN-GLOBAL and KNN-INSIDER satellite models across all ecoregions.	130
Figure S2.47. Performance (% bias) of RF-GLOBAL and RF-INSIDER lidar models across all ecoregions.	131
Figure S2.48. Performance (% bias) of RF-GLOBAL and RF-INSIDER satellite models across all ecoregions.	132
Figure 3.1. FVS Variants and Locations across CONUS	135
Figure 3.2. Hierarchical models as a partial pooling of parameters across related groups	141
Figure 3.3. Layout of FVS Variants, Locations, and FIA Plot Data.....	145
Figure 3.4. FIA plot layout with multiple conditions mapped across subplots	146
Figure 3.5. Illustration of the Bayesian State-Space model.....	153
Figure 3.6. Adjustment to DBH measurement error.....	155
Figure 3.7. Illustration of HDI-ROPE decisions.....	160
Figure 3.8. Unpooled model comparison with multiple performance metrics: Douglas-fir	164
Figure 3.9. HDI-ROPE comparisons between FVS default models and hierarchical benchmark models, Part 1 of 2.	167

Figure 3.10. HDI-ROPE comparisons between FVS default models and hierarchical benchmark models, Part 2 of 2.	168
Figure 3.11. Posterior distributions of parameters for the hierarchical benchmark model for Douglas-fir.	175
Figure S3.12. Unpooled model comparison with multiple performance metrics: Ponderosa pine.	184
Figure S3.13. Unpooled model comparison with multiple performance metrics: Western hemlock.	185
Figure S3.14. Unpooled model comparison with multiple performance metrics: White fir	186
Figure S3.15. Unpooled model comparison with multiple performance metrics: Pacific silver fir.	187
Figure S3.16. Unpooled model comparison with multiple performance metrics: Mountain hemlock.	188
Figure S3.17. Unpooled model comparison with multiple performance metrics: Western redcedar.	189
Figure S3.18. Unpooled model comparison with multiple performance metrics: Canyon live oak.	190
Figure S3.19. Unpooled model comparison with multiple performance metrics: Tanoak.	191
Figure S3.20. Unpooled model comparison with multiple performance metrics: Red alder.	192
Figure S3.21. Unpooled model comparison with multiple performance metrics: California black oak.	193
Figure S3.22. Unpooled model comparison with multiple performance metrics: Pacific madrone.	194
Figure S3.23. Unpooled model comparison with multiple performance metrics: Coast redwood.	195
Figure S3.24. Unpooled model comparison with multiple performance metrics: Bigleaf maple.	196
Figure S3.25. Unpooled model comparison with multiple performance metrics: Noble fir.	197
Figure S3.26. Posterior distributions of parameters for the hierarchical benchmark model for Ponderosa pine.	198

Figure S3.27. Posterior distributions of parameters for the hierarchical benchmark model for Western hemlock.	199
Figure S3.28. Posterior distributions of parameters for the hierarchical benchmark model for White fir.	200
Figure S3.29. Posterior distributions of parameters for the hierarchical benchmark model for Pacific silver fir.	201
Figure S3.30. Posterior distributions of parameters for the hierarchical benchmark model for Mountain hemlock.	202
Figure S3.31. Posterior distributions of parameters for the hierarchical benchmark model for Western redcedar.	203
Figure S3.32. Posterior distributions of parameters for the hierarchical benchmark model for Canyon live oak.	204
Figure S3.33. Posterior distributions of parameters for the hierarchical benchmark model for Tanoak.	205
Figure S3.34. Posterior distributions of parameters for the hierarchical benchmark model for Red alder.	206
Figure S3.35. Posterior distributions of parameters for the hierarchical benchmark model for California black oak.	207
Figure S3.36. Posterior distributions of parameters for the hierarchical benchmark model for Pacific madrone.	208
Figure S3.37. Posterior distributions of parameters for the hierarchical benchmark model for Coast redwood.	209
Figure S3.38. Posterior distributions of parameters for the hierarchical benchmark model for Bigleaf maple.	210
Figure S3.39. Posterior distributions of parameters for the hierarchical benchmark model for Noble fir.	211

LIST OF TABLES

Table 1.1. Silvicultural systems ¹ and treatments applied to each of the parcels.	13
Table 1.2. Delivered log prices.	17
Table 1.3. Management costs.	17
Table 1.4. Isolated impacts of green tree retention and RMZ widths constraints on KPIs for each scenario.	26
Table 1.5. Additional value for wood and carbon storage required to close the financial gap of alternative management scenarios with the BAU scenario.....	26
Table S1.6. FVS keyword modifications.....	35
Table S1.7. Mean percent defect incorporated into FVS for each species diameter at breast height	37
Table S1.8. Yield Table: Dominant Height (feet), 40 tallest trees per acre.....	44
Table S1.9. Yield Table: Trees per Acre	44
Table S1.10. Yield Table: Basal Area (sq. ft. per acre).....	45
Table S1.11. Yield Table: Quadratic Mean Diameter (inches)	45
Table S1.12. Yield Table: Cubic volume including top and stump (hundreds of cubic feet per acre).....	46
Table S1.13. Yield Table: Boardfoot Volume, Scribner Rule (thousand boardfeet per acre) *	46
Table S1.14. Yield Table: Stand Density Index	47
Table S1.15. Yield Table: Curtis Relative Density.....	47
Table S1.16. Buffer widths for Riparian Management Zones under FPA and FSC Rules (in feet)	48
Table S1.17. RMZ length for 50-foot core (no-touch) buffer area for Type Np streams that intersect with Type S or Type F streams, Washington State.	49
Table S1.18. Minimum Retention for Inner RMZ Buffers on Type F streams, Washington State	50

Table S1.19. Oregon FPA “Standard Targets” for Retention in Inner RMZ Buffers.....	51
Table S1.20. Oregon FPA “Alternative Targets” for Retention in Inner RMZ Buffers...	52
Table 2.1. Studies investigating the transferability of forest structure models trained with remote sensing data	56
Table 2.2 Plot-level observations used for model training and evaluation from each year of lidar acquisitions and satellite imagery across Pacific Northwest ecoregions.	63
Table 2.3. Predictor variables used in lidar-based and satellite-based predictive models.	66
Table 2.4. Time lags between field observations and remote sensing.....	68
Table 2.5. Practical equivalence between Global and Insider models, NRMSE	96
Table 2.6. Practical equivalence between Global and Insider models, % Bias	97
Table 3.1. Samples sizes of trees used to fit current FVS models and from FIA data used for refitting models in this study.....	140
Table 3.2. Repeatedly measured trees used for model training and evaluation.....	148
Table 3.3. Parameterization of the State-Space Models	157
Table 3.4. Summary of Unpooled Model Comparisons	162
Table 3.5. Performance metrics for FVS and refitted models: Hardwood species.....	165
Table 3.6. Performance metrics for FVS and refitted models: Softwood species	166
Table 3.7. Distributions of standardized model coefficients for hierarchical SIMPLE-BALLNDBH-BA models	169
Table 3.8. Mean values of unstandardized model coefficients for hierarchical SIMPLE-BALLNDBH-BA models	172
Table 3.9. Standard deviations of varying intercepts and random effects for the hierarchical SIMPLE-BALLNDBH-BA model	176
Table S3.10. Prioritization of FVS research and development topics by the Forest Carbon Modeling Group.....	182
Table S3.11. Hierarchical parameterization of Chapman-Richards height-diameter model	183

ACKNOWLEDGEMENTS

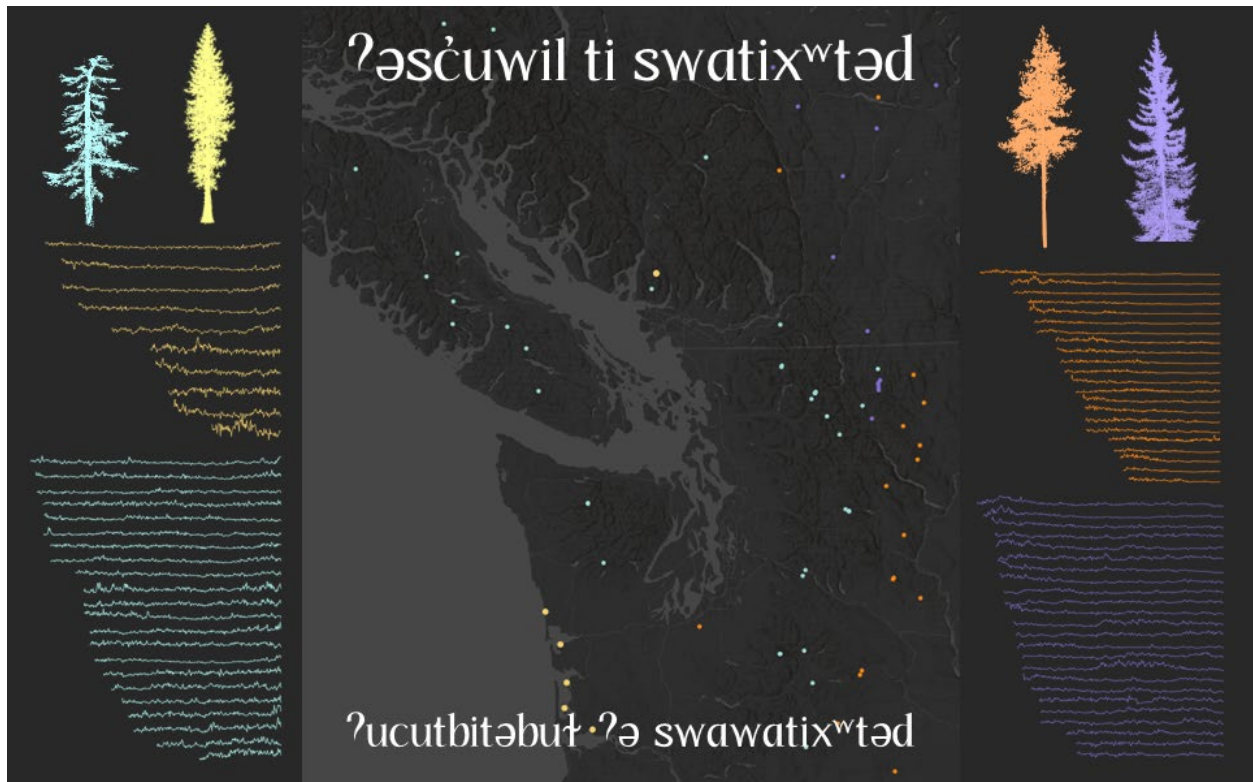
This effort has been one of the most challenging in my life. Over the past seven years, I have been advised and cared for by more people than I can name. I am particularly grateful to my supervisors at Ecotrust and the University of Washington, Brent Davies and Gregory Ettl, to my teammates at Ecotrust Sara Loreno and Stephanie Gutierrez, and to my wife, Abby Cullinan. Each of these people helped carry what I could not and supported me in innumerable ways to be able to provide care for my family as my highest priority. I have grown immensely, both personally and professionally, thanks to these people in particular. I could not have made it to this point without any of them.

My work and research has been supported financially by several individual donors, state and federal agencies, and foundations, including: the University of Washington School of Environmental and Forest Sciences; the University of Washington College of the Environment; the Northwest Climate Adaptation Science Center; the ARCS Foundation; the Bullitt Foundation; the MillsDavis Foundation; the ZG Foundation; the Weyerhaeuser Family Foundation; the USDA Natural Resources Conservation Service and USDA Oregon Natural Resources Conservation Service; the US Forest Service PNW Research Station; the USDA Western Sustainable Agricultural Research & Education (SARE) program, and by the Oregon Tree Farm System.

I have benefited greatly from collaboration with several partner organizations and inspiring leaders and advisors from Northwest Natural Resources Group, Wallowa Resources, EFMI, Washington Environmental Council, Oregon State University Extension, Oregon Tree Farm System, the Hoopa Valley Tribe, the Swinomish Tribe, and the Confederated Tribes of Coos, Lower Umpqua and Siuslaw Indians. In particular, I would also like to acknowledge the Advisory Team for a project funded by USDA Western SARE: Colin Beck; Nate Agalzoff; Cherie Kearney; Lynn Baker; Ben Hayes; Andrew Owen; George McKinley; and Lauren Grand. Thank you.

DEDICATION

For the tall people, so that walking people may better know and honor them.



“Trees tell us the world is beautiful.” (David Diaz, 2019)

Chapter 1. Tradeoffs in Timber, Carbon, and Cash Flow under Alternative Management Systems for Douglas-Fir in the Pacific Northwest

This chapter is reproduced from an article that has been published in the journal *Forests*. The copyright is held by the authors under the Creative Commons Attribution 4.0 International License (CC BY 4.0). This License permits free sharing, copying, and redistribution of the work as long as the original source is attributed and any changes that have been made are indicated without implying endorsement by the authors. A full copy of the License can be found here: <https://creativecommons.org/licenses/by/4.0/>. The preferred citation for this article is:

Diaz, D.D.; Loreno, S.; Ettl, G.J.; Davies, B. Tradeoffs in Timber, Carbon, and Cash Flow under Alternative Management Systems for Douglas-Fir in the Pacific Northwest. *Forests* **2018**, *9*, 447.
<https://doi.org/10.3390/f9080447>

1.0 ABSTRACT

Forest management choices offer significant potential to mitigate global climate change and biodiversity loss. To illuminate tradeoffs relevant to policymakers, forest sector stakeholders, and consumers of forest products, we utilize three Key Performance Indicators—average carbon storage in the forest and wood products; cumulative timber output; and discounted cash flow—to compare four alternative management scenarios for Douglas-fir forests on 64 parcels across western Oregon and Washington. These scenarios are designed to meet one of two alternative management objectives: (i) maximize Net Present Value; or (ii) maximize sustained timber yield; according to one of two alternative sets of forest practice constraints: (i) compliance with minimum Oregon/Washington Forest Practices Act (FPA) rules; or (ii) two key requirements (increased green tree retention and wider riparian buffers) of Forest Stewardship Council (FSC) certification. Improved performance in terms of carbon storage for these alternatives generally also corresponded with reduced Net Present Value and timber yields. The gap between FSC and FPA performance indicators was wider in Oregon than Washington, which is primarily attributed to the higher level of stream protection required under Washington versus Oregon FPA rules. We

observed consistently higher average carbon storage per cumulative timber output among FSC scenarios relative to business-as-usual, indicating FSC-certified wood carries an embedded carbon benefit. Our findings highlight options for targeted policies to incentivize management that increases carbon storage and minimizes disruptions in timber output, as well as for narrowing the financial gap (or opportunity cost) that would be involved in a transition away from contemporary common practice on industrial timberlands in the coastal Douglas-fir forests of the Pacific Northwest.

1.1 INTRODUCTION

The Productivity and Management of Coastal Douglas-Fir Forests

The coastal temperate forests of the Pacific Northwest are among the most productive ecosystems in the world (Waring and Franklin, 1979). Long-lived tree species such as Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco), western hemlock (*Tsuga heterophylla* (Raf.) Sarg.), and western redcedar (*Thuja plicata* Don ex D. Don) and conducive growing conditions (Gholz, 1982) interact to enable the accumulation of immense forest biomass over time (Franklin and Dyrness, 1988). Oregon and Washington are the two largest producers of softwood lumber in the USA, generating 16.5% and 11.8% of the country's softwood lumber supply in 2015 (Oregon Forest Resources Institute, 2017). Roughly half of the productive forestland in the region, and nearly 90% of forest industry softwood timberland, is covered by Douglas-fir and western hemlock forest types (Haynes, 2003). Douglas-fir dominates timber volume and revenue production in the region and generally plays a foundational role in management, conservation, and policy decisions related to forests throughout the Pacific Northwest.

The management of Douglas-fir on industrial timberland in the Pacific Northwest west of the Cascade Range (see project area in Figure 1.1) is generally intensive, following an even-age silvicultural system including the selection of genetically superior planting stock, site preparation and the broadcast application of herbicides to limit shrub and broadleaf competition, and regeneration harvests (i.e., harvest to initiate a new stand) around financially optimal rotation ages of 40–50 years (Haynes, 2003; van Kooten et al., 1995; Talbert and Marshall, 2005).

The restructuring of industrial timberland ownership and vertical disintegration of major timber industry firms in the USA since the 1980s has given rise to an increasingly consolidated set of “financialized” investment entities such as Timberland Investment Management Organizations and Real Estate Investment Trusts (Binkley, 2007; Gunnoe, 2016; Bliss et al., 2010). These new timberland brokers and owners, backed primarily by institutional investors, have since replaced nearly all the largest industrial private forest companies in the USA (Bliss et al., 2010). This ownership shift has been described as the “financialization of landownership” and corresponds with an emphasis on silvicultural systems designed to maximize the return on investment from timberland (Binkley, 2007; Bliss et al., 2010).

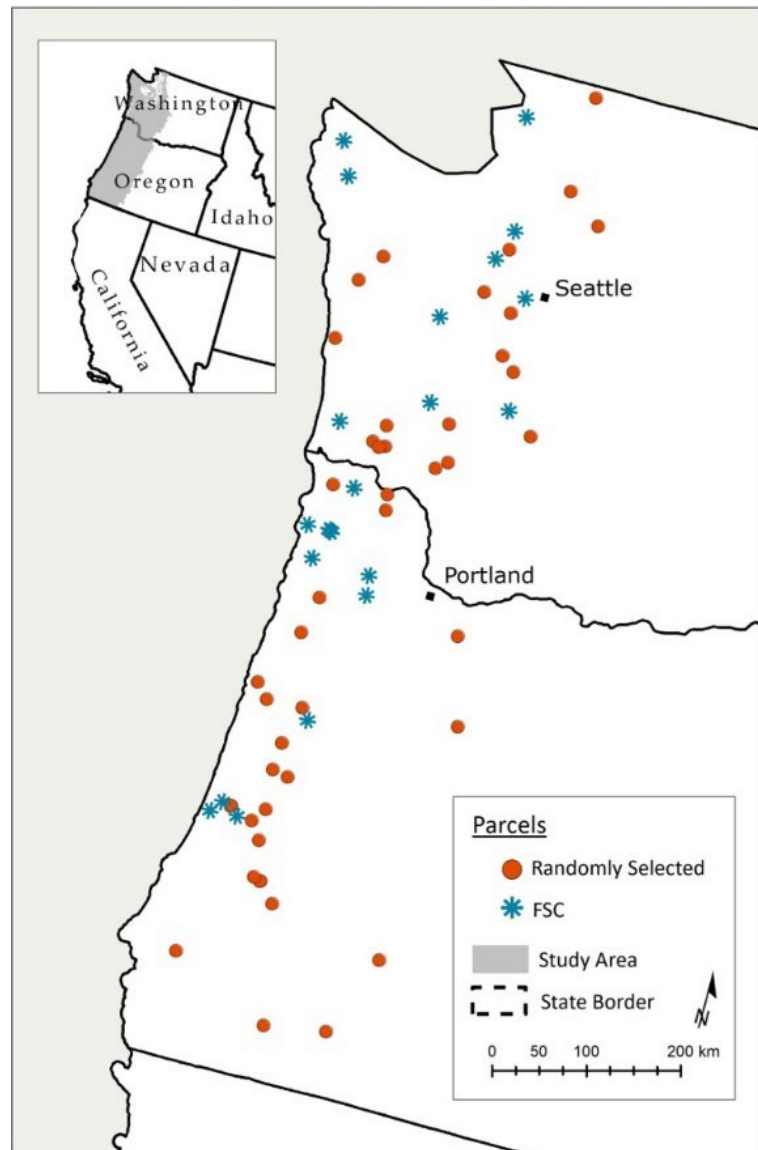


Figure 1.1. Study area.

Actual FSC-certified parcels are distinguished from randomly selected parcels. All parcels are subject to the same suite of simulated management practices (both ~FSC and ~FPA scenarios). FPA: Forest Practices Act; FSC: Forest Stewardship Council.

The practice of managing Douglas-fir forests in the Pacific Northwest for maximum sustained yield of timber (i.e., rotations that maximize timber yield vs. those that maximize return on investment) has become progressively less common among private timberland owners (Binkley, 2007). For even-aged silvicultural systems, the divergence between the ‘biological rotation age’ at which maximum sustained yield would be achieved and younger ‘financial rotation

age’ at which maximum return on investment would be achieved has been well-recognized in forestry since the nineteenth century, and has played a fundamental role in forest economics ever since (Brazee, 2001).

A variety of long-term field trials and simulation studies have demonstrated an opportunity for long-term timber supply to be increased by the extension of rotation ages closer to the culmination of Mean Annual Increment (MAI) of timber volume growth, and that the integration of commercial thinning may extend the culmination of MAI further into substantially older ages than current common practice (Brazee, 2001; Curtis, 1995; Curtis and Marshall, 1993; Curtis, 2006, 1997; Curtis et al., 2004). The extension of rotation ages is also well-established as a means to increase average carbon storage in the forest system and is a common component in forest carbon policies and research (van Kooten et al., 1995; Foley et al., 2009; Sohngen and Brown, 2008; Law et al., 2018; Tóth et al., 2013; Fischer et al., 2017; Harmon and Marks, 2002). Previous work investigating the potential for carbon sequestration incentives to support the extension of rotation ages have specifically identified the coastal forests of the Pacific Northwest possessing unrivaled sequestration potential under a variety of carbon accounting approaches (Foley et al., 2009).

Policy Interest in Forest Sector Engagement in Climate Change Mitigation and Adaptation

In recent years, climate policy proposals in both Oregon and Washington have considered forests for potential involvement in climate change mitigation as well as for investments of climate program funds for adaptation. Although no economy-wide regulations of greenhouse gas emissions have yet been implemented, State-level policy proposals based on a carbon tax or fee, cap-and-trade, or variations such as “cap-and-invest” or “cap-and-dividend” have become an almost-yearly occurrence and are currently active in both Oregon and Washington via the State legislature and through citizens’ ballot initiatives. The role of forests within these programs is especially relevant to local policymakers and forest sector stakeholders due to the historical natural resource dependence of many rural communities in the Pacific Northwest and the growing awareness of climate impacts on forests and communities in the region.

In this study, we focus primarily on the potential for forest management to contribute to the mitigation side of this policy discussion because we anticipate incentive programs and potential regulatory approaches dealing with the impacts of alternative forest management approaches on

carbon balance. The exceptional productivity and biomass carrying capacity of coastal forests in the Pacific Northwest justifies the attention paid to them, both in terms of private and public capital investments in timberland management, as well as consideration of the climate change mitigation potential these forests may offer.

Interest in the Carbon Footprint of Wood, and the Central Role Certification Has Come to Play

A growing emphasis on green building and reducing environmental impacts in architecture and engineering fields has led to a proliferation of Environmental Product Declarations and a corresponding interest in wood as preferable building material relative to more energy- and carbon-intensive alternatives (Sathre and O'Connor, 2010; Bergman and Taylor, 2011). Our original motivation for this work evolved from an analysis of ecosystem service impacts involved in the construction and maintenance of The Bullitt Center, a six-story office building in Seattle, Washington, USA designed to meet the standards developed by the International Living Future Institute: The Living Building Challenge (Cowan et al., 2014). As part of this analysis, we were tasked to quantify the embedded carbon storage impact involved with the procurement of wood certified according to the Forest Stewardship Council (FSC) program. We evaluated the “upstream” carbon implications of forest management as a distinct process from the “downstream” design decisions regarding the use of wood or other materials in the construction of the building.

Programs that certify the sustainable management of forests have grown to assume a fundamental role as independent gatekeepers for emerging programs focused on reducing the environmental impacts humans produce. For example, the Living Building Challenge requires all wood used in projects to be FSC certified, from salvaged sources, or intentionally harvested from onsite. A similar preference existed for FSC-certified wood in the Leadership in Energy and Environmental Design (LEED) green building certification for many years. Third-party certification has also become a virtually universal requirement for all major forest carbon offset crediting programs now in operation in the USA and abroad, with an estimated 99.7% of carbon credits transacted globally involving the use of third-party certification (Hamrick and Gallant, 2017). The most widely subscribed forest carbon program in the USA is operated under the cap-and-trade program maintained by the State of California's Air Resources Board, which requires forest projects to demonstrate an independent certification of sustainable forest management in addition to the carbon offset verification in order to be eligible for carbon crediting.

Despite the preference for FSC in green building programs, very few examples exist of research quantifying the impacts that the additional constraints on forest practices required to achieve FSC certification may have. The most common approach to this type of analysis has focused exclusively on traditional forestry indicators such as timber output, forest sector employment in forest products processing, or financial performance of individual ownerships (Mendell and Lang, 2013; Eriksson et al., 2007; Nebel et al., 2005; Bouslah et al., 2010), ignoring the broader set of values, including ecological and social impacts, that form the basis of FSC Principles and Criteria for certification (FSC International, 2015). The integrated consideration of environmental impacts corresponding with FSC certification remains relatively sparsely covered in the peer-reviewed literature (Tóth et al., 2013). In this study, we explore these impacts and the potential for FSC certification to function as a surrogate for more elaborate carbon offset crediting mechanisms.

Research Objectives and Working Hypotheses

Our primary objective in this study is to quantify the impacts that selected silvicultural practices associated with FSC-certification have compared to business-as-usual forest management approaches. We build on earlier work by integrating both environmental and economic indicators that are fundamental concerns in ongoing discussions of forest carbon and climate policy. We apply three Key Performance Indicators (KPIs)—carbon storage, timber output, and discounted cash flow—to illustrate the potential for new policies such as forest carbon incentives to reduce financial barriers to the broader adoption of forest practices that increase climate change mitigation and a host of other ecosystem services.

In general, we expect silvicultural systems which employ longer rotations (timed to the culmination of Mean Annual Increment of merchantable timber volume) to carry and yield larger timber volumes and store more carbon over time than the more financially attractive shorter rotations. While greater timber output and carbon storage through extended rotations may seem like a clear win-win from a policy perspective, we expect gains in carbon storage and sustained timber yield to come at the expense of a non-trivial financial gap (or opportunity cost). This financial gap is likely to present a substantial barrier to the adoption of new forest practices by public and private landowners that could be engaged through programs focused on sequestering more carbon and/or generating more timber (Rabotyagov and Lin, 2013).

We also expect forest practices that retain more trees during regeneration harvests and which reduce or exclude intensive management around streams to translate into higher forest carbon storage over time. Forest certification programs that impose these types of constraints on harvest practices, such as the FSC-US Standard in the Pacific Northwest, are therefore likely to translate into additional carbon storage in the forest with a corresponding reduction in timber yield and Net Present Value.

1.2 MATERIALS AND METHODS

Initial Forest Conditions

We assess the impact of alternative silvicultural practices using a range of forest conditions across landscapes where intensive Douglas-fir management is commonly practiced in the Pacific Northwest. A total of 64 parcels (covering 44,250 acres) were selected across western Oregon and Washington with the intent of covering a spectrum from small-to-large parcel size, as well as from sparse-to-dense coverage of riparian areas. Twenty-two of these parcels (covering 10,319 hectares/25,500 acres) have been FSC-certified (Figure 1.1). We selected the remaining 42 parcels (covering 7,515 hectares/18,570 acres) from privately owned forest parcels larger than 40.5 hectares (100 acres) within the study region. These parcels were selected to offer a diverse set of parcel sizes and extent of riparian cover and are intended to communicate the range of variability that might be expected through the type of management scenarios we consider in this study if they were applied more broadly across the region.

Because ground-based inventory data were not available across all parcels to delineate stands or provide starting conditions, we utilized remotely sensed data to estimate initial forest composition. We delineated each parcel using a 2.02 hectare (5 acre) hexagonal grid to approximate management units, which are hereafter referred to as stands (Figure 1.2). Forest inventory data were imputed to each stand from the Landscape Ecology, Modeling, Mapping & Analysis group's Gradient Nearest Neighbor (GNN) database (Ohmann and Gregory, 2002; Ohmann et al., 2011; LEMMA, 2014). The GNN database provides a crosswalk between a raster image with 30×30 m resolution to forest inventory plots for each pixel. We summarized pixel-level inventory estimates up to the scale of the hexagonal stand using the majority forest type across pixels within each hexagon and selected the most commonly identified inventory plot of

that forest type within the hexagon to represent that stand (i.e., a single forest inventory plot of the

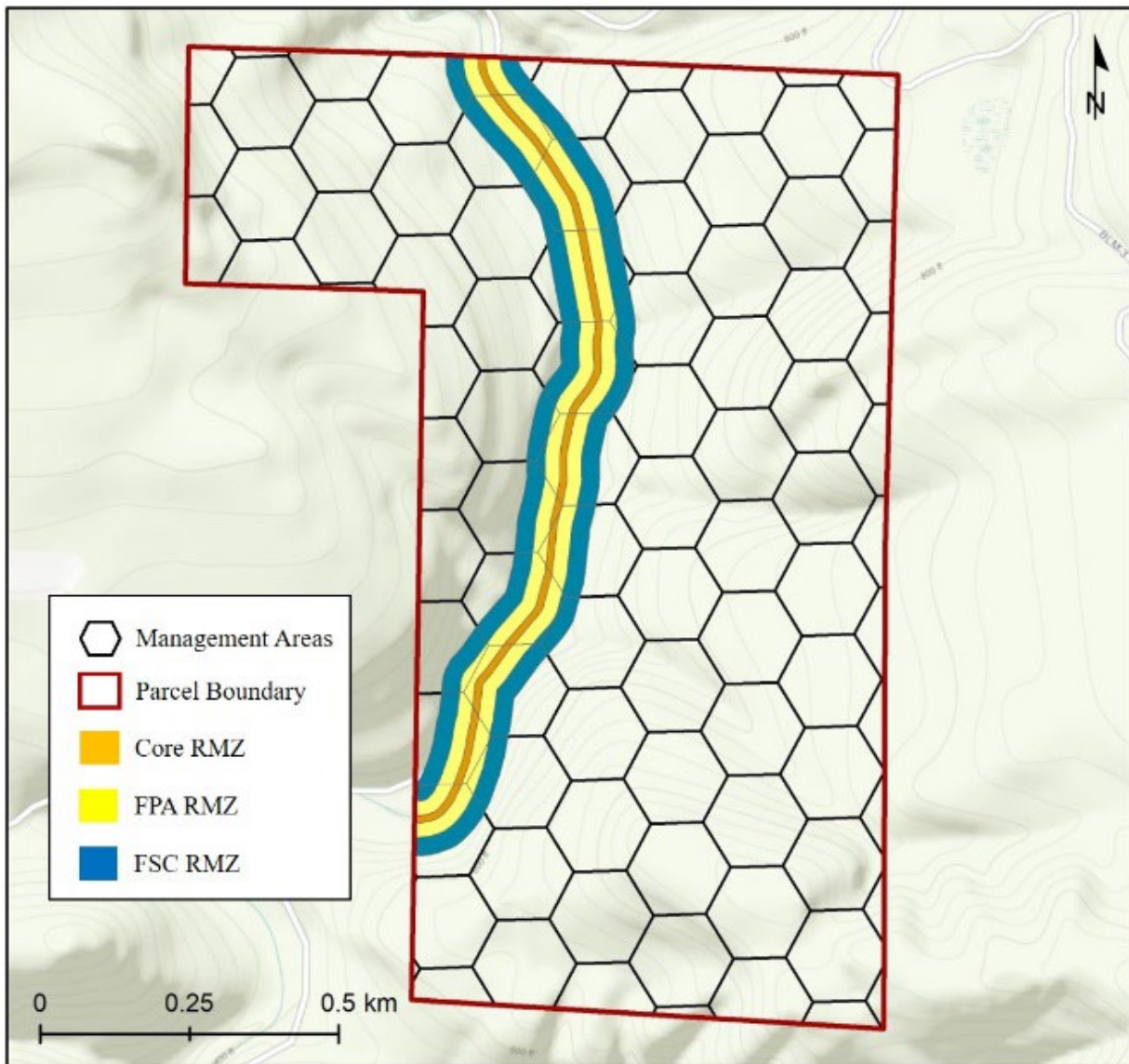


Figure 1.2. Division of landscape into management units to be simulated.

Each parcel is sub-divided into 2.02 hectare (5 acre) 'stands', which are then intersected with Riparian Management Zones (RMZs) to form distinct management units that are simulated individually. The parcel above shows the core and non-core RMZ areas delineated following FSC rules.

majority forest type to represent each stand).

We imputed topographical attributes including elevation, aspect, and slope to each stand based on a Digital Elevation Model. Douglas-fir 50-year Site Index was estimated at the stand level based on maps produced by Latta et al. (Latta et al., 2009). These Site Index predictions of

Latta et al. were based on climatic drivers and vary relatively smoothly across the landscape; they are therefore unlikely to accurately reflect changes in site productivity following topographic or edaphic changes at small spatial scales (e.g., along streams or in rocky outcrops). Several measures illustrating the correspondence between the imputed attributes of the parcels selected for modeling and the landscape from which they were sampled are shown in Figure 1.3. Both the randomly selected and actual FSC-certified parcels are generally representative of the landscape across which they occur. The starting inventory conditions for randomly selected and FSC-certified parcels in terms of stand age, basal area, volume, and biomass, are comparable in both Oregon and Washington.

Management Systems

In this study, we consider even-aged wood-production-oriented silvicultural systems for Douglas-fir. We do not evaluate other single species systems (e.g., red alder (*Alnus rubra* Bong.)), those which intentionally retain more than one species (e.g., mixtures of Douglas-fir and western hemlock that are also common in western Oregon and Washington), or any uneven-aged management systems in this study. For each parcel, we develop four alternative management scenarios which represent the unique combinations of two alternative objectives and two alternative sets of management constraints. These four management scenarios will be referred to as “BAU” (for Business as Usual), “SHORT~FSC”, “LONG~FSC”, and “LONG~FPA”. These are each described in greater detail below.

Management Objectives

We designed our management scenarios to achieve one of two alternative management objectives: maximize Net Present Value (NPV), or maximize the sustained yield of timber (in terms of cubic volume of sawlogs). For each of these two objectives, we determined an optimal rotation age. “Financially optimal” rotation ages for the NPV-maximizing scenarios were identified as the age at which NPV peaked as the forest grew over time. For each financially optimal rotation, the Soil Expectation Value (SEV), or the present value of perpetual management of the land using that rotation age was also calculated. We calculated “biologically optimal” rotation ages for yield-maximizing scenarios as the age at which the Mean Annual Increment of total cubic volume culminated.

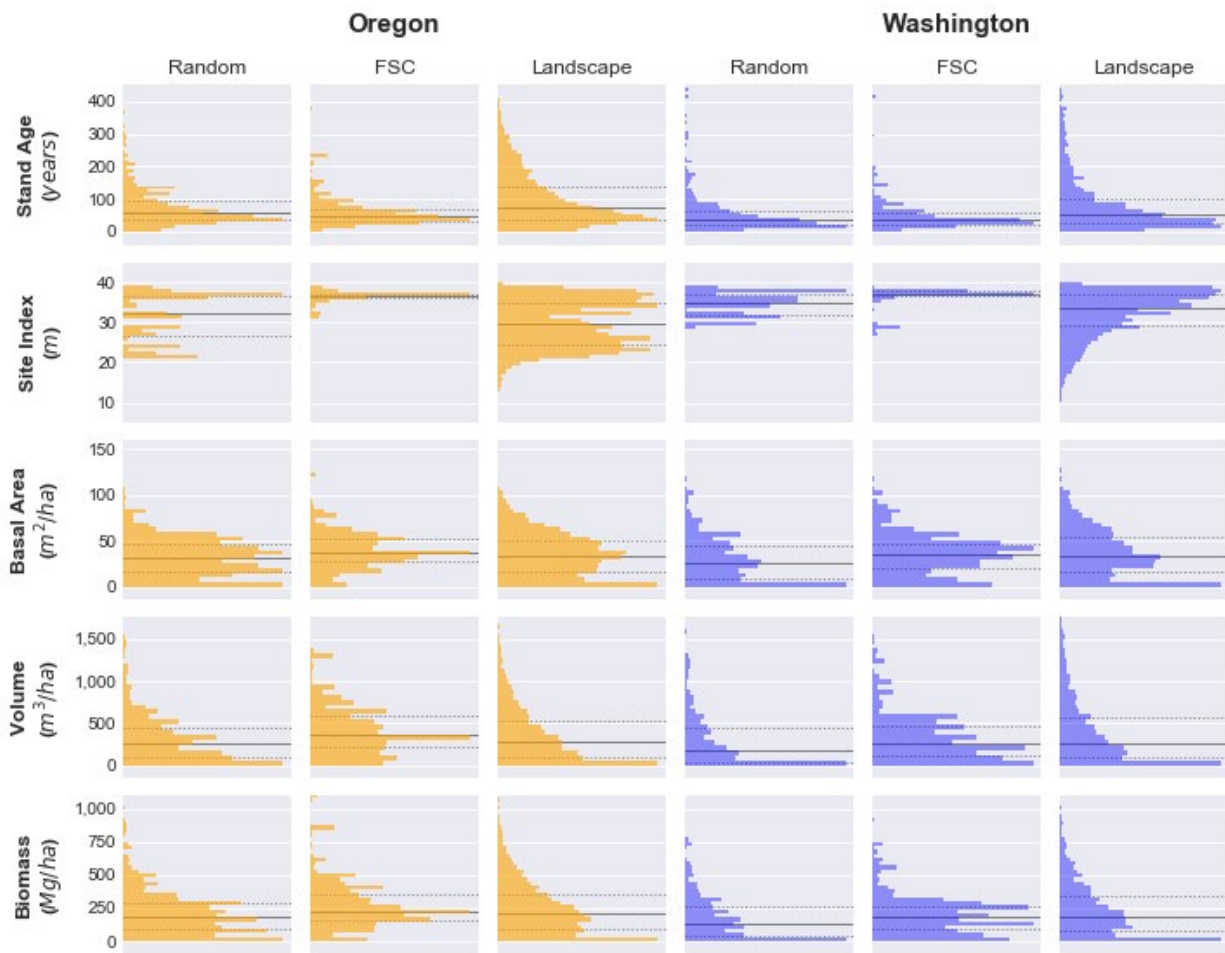


Figure 1.3. Correspondence between selected parcels and the surrounding landscape. The “surrounding landscape” refers to the extent of the US Environmental Protection Agency Level 3 Ecoregions in western Oregon and Washington which include the sampled parcels. Only areas with forest landcover according to the Gradient Nearest Neighbor Plot Database for 2014 (LEMMA, 2014) are included in the calculation of these distributions, which are visualized with a histogram based on counts of 30×30 -meter pixels within each area of interest. Horizontal lines in each graph showing the 25th and 75th percentiles (dotted lines) and median (solid line). Site Index refers to the height of dominant Douglas-fir trees at age 50, with values derived from Latta et al. (2009).

For both objectives, existing stands—with inventory derived from the GNN database—are converted into a Douglas-fir forest based on a financially optimal conversion time determined following Martin (Martin, 2013). Briefly, the optimal conversion time for each stand was calculated when ‘value of forest’ peaks starting with current forest conditions. As described by Martin (2013), ‘value of forest’ is calculated as the sum of the NPV of harvestable timber (‘value of trees’) in each year plus the discounted SEV for the future management of the stand as a

Douglas-fir plantation. The initial regeneration/conversion harvest honored the green tree retention and RMZ constraints described below (and in Supplementary Materials) and was followed by planting of Douglas-fir. The varying levels of retention in each scenario produce different forest structure and residual species composition after this initial harvest and over time, although they all move the stands increasingly towards Douglas-fir monocultures across the 100-year simulation timeframe.

Management Constraints

We impose two sets of alternative management constraints for each management objective related to two primary silvicultural choices: the retention of live trees during regeneration harvests; and the limitation or prohibition of harvest activities within Riparian Management Zones (RMZs). The first set of constraints we model represents compliance with the minimum requirements of the Oregon and Washington Forest Practices Acts (FPA) (Washington State Legislature, n.d.; Oregon Secretary of State, n.d.). The second set of constraints represents compliance with two of the primary requirements for certification under the Forest Stewardship Council (FSC) program (FSC-US, 2012).

Although we model these constraints distinctly for each State, we do not capture the full suite of requirements for compliance with either FPA or FSC rules. FSC tends to impose substantial restrictions above and beyond the FPA rules in both States, although the additional requirements for FSC certification are generally more pronounced for Oregon than for Washington, which is primarily due to the difference in limitations on harvesting near streams between Oregon and Washington FPA rules (see Figure 1.4). To help distinguish our selective choice of management constraints from the full FSC and FPA rules, we denote the management scenarios we simulate as ~FSC and ~FPA, respectively, reserving the unadorned acronyms FSC and FPA to refer to actual FSC-certified parcels or the FSC standard and to FPA rules and regulations.

We also do not apply the full flexibility permitted within FPA or FSC rules. For example, we do not simulate the removal of trees left to satisfy green tree retention in non-riparian stands at the time of a harvest during subsequent entries in neighboring management units, although this appears to be common practice in Oregon and Washington. In our treatment of RMZs, retention requirements are satisfied within each RMZ-designated polygon as a standalone management unit.

Although FSC and FPA rules in both Oregon and Washington permit it, we do not count retention in RMZs towards retention requirements in adjacent non-RMZ harvest units. Oregon and Washington FPA rules assign varying RMZ widths and permit varying intensities of harvest within RMZs depending on the stream type (e.g., fish- or non-fish-bearing perennial/seasonal), stream size (Washington and Oregon), and site class (Washington). Although FSC permits single-tree harvest in inner RMZ buffers and group selection in outer RMZ buffers, we have modeled all FSC RMZs as no-touch because neither single-tree nor group selection harvests have been simulated. In this context, the management actions simulated in this study are likely to portray a reduced intensity of management in RMZs and adjacent harvest units than are technically permissible under FPA and FSC rules.

Our simulation of ~FPA management scenarios generally errs on the side of conserving more trees during harvests than would be likely following minimal compliance with FPA rules. This is related in part to our choice not to model a variety of exceptions to general RMZ protections (e.g., allowing the successive conversion of hardwood-dominated riparian stands through harvest and establishment of new conifer cover over time).

Our simulation of ~FSC management scenarios also leaves out several important requirements of FSC certification. For example, FSC rules discourage and limit the conversion of natural forests to plantations, which we simulate in our study. FSC certification requires at least 10–25% of a Forest Management Unit to be maintained and/or restored to a natural or semi-natural state wherever natural ecosystems have been previously converted to plantations. We do not incorporate these constraints in our simulations but recognize that other studies considering set-aside and maximum contiguous harvest area restrictions in FSC have indicated these constraints would lead to additional reductions in harvestable timber volume over time beyond the impacts of those constraints we included in this study (Tóth et al., 2013; Mendell and Lang, 2013).

Silvicultural Systems

The combination of the two objectives (NPV or MSY) and two management constraints (FPA or FSC) produce four separate management scenarios for each parcel, which are elaborated in Table 1.:

1. **BAU or SHORT~FPA.** “Business-as-usual” (BAU) management to maximize NPV under the selected constraints of State FPA rules. This scenario represents common practice in production forests of western Oregon and Washington.
2. **SHORT~FSC.** Management to maximize NPV under the green tree retention and RMZ constraints required by FSC. “Short” refers to the relative length of the rotation age compared to MSY-oriented scenarios.
3. **LONG~FPA.** Management to maximize the sustained yield of timber under the selected constraints of State FPA rules. “Long” refers to the relative length of the rotation age relative to the NPV-based management scenarios.
4. **LONG~FSC.** Management to maximize the sustained yield of timber under the selected constraints of FSC certification.

Table 1.1. Silvicultural systems¹ and treatments applied to each of the parcels.

Activity	BAU	SHORT~FSC ¹	LONG~FPA	LONG~FSC ¹
Planting Douglas-fir	1,075 tph (435 tpa)	1,075 tph (435 tpa)	1,075 tph (435 tpa)	1,075 tph (435 tpa)
Commercial Thinning ²	None	None	@ 55% SDImax ³ , thin to 45% SDImax	@ 55% SDImax, thin to 45% SDImax
	@ 38–44 years	@ 38–44 years	@ 75 years	@ 75 years
Regeneration Harvest	retain 10 tph ≥ 30.5 cm DBH (4 tpa ≥ 12 in DBH)	retain 30% pre-harvest basal area	retain 10 tph ≥ 30.5 cm DBH (4 tpa ≥ 12 in DBH)	retain 10% of pre- harvest basal area

¹ Although these silvicultural systems are identified as ~FSC scenarios, they are not intended to reflect the silvicultural systems now practiced by existing FSC-certified landowners. In general, silvicultural systems we simulated are more intensive than those practiced by FSC-certified landowners in the Pacific Northwest.

² Commercial thinning was allowed after age 30, with re-entry as frequently as every 15-years if at least 7 MBF/ha (3 MBF/ac) in harvest volume would be generated and only if the regeneration harvest was not scheduled for at least another 15 years.

³ Methods for determining maximum Stand Density Index (SDImax) are described briefly below and in the Supplementary Materials. DBH: diameter at breast height.

Key Performance Indicators

In this study, we utilize three Key Performance Indicators (KPIs) to compare forest management scenarios: carbon storage, timber output, and discounted cash flow. These KPIs provide information about the potential direct tradeoffs between a newly incentivized ecosystem service, carbon storage, and more traditional KPIs from the forest sector including timber output (which is often used to calculate down-stream economic/job impacts) and cash flow to the landowner.

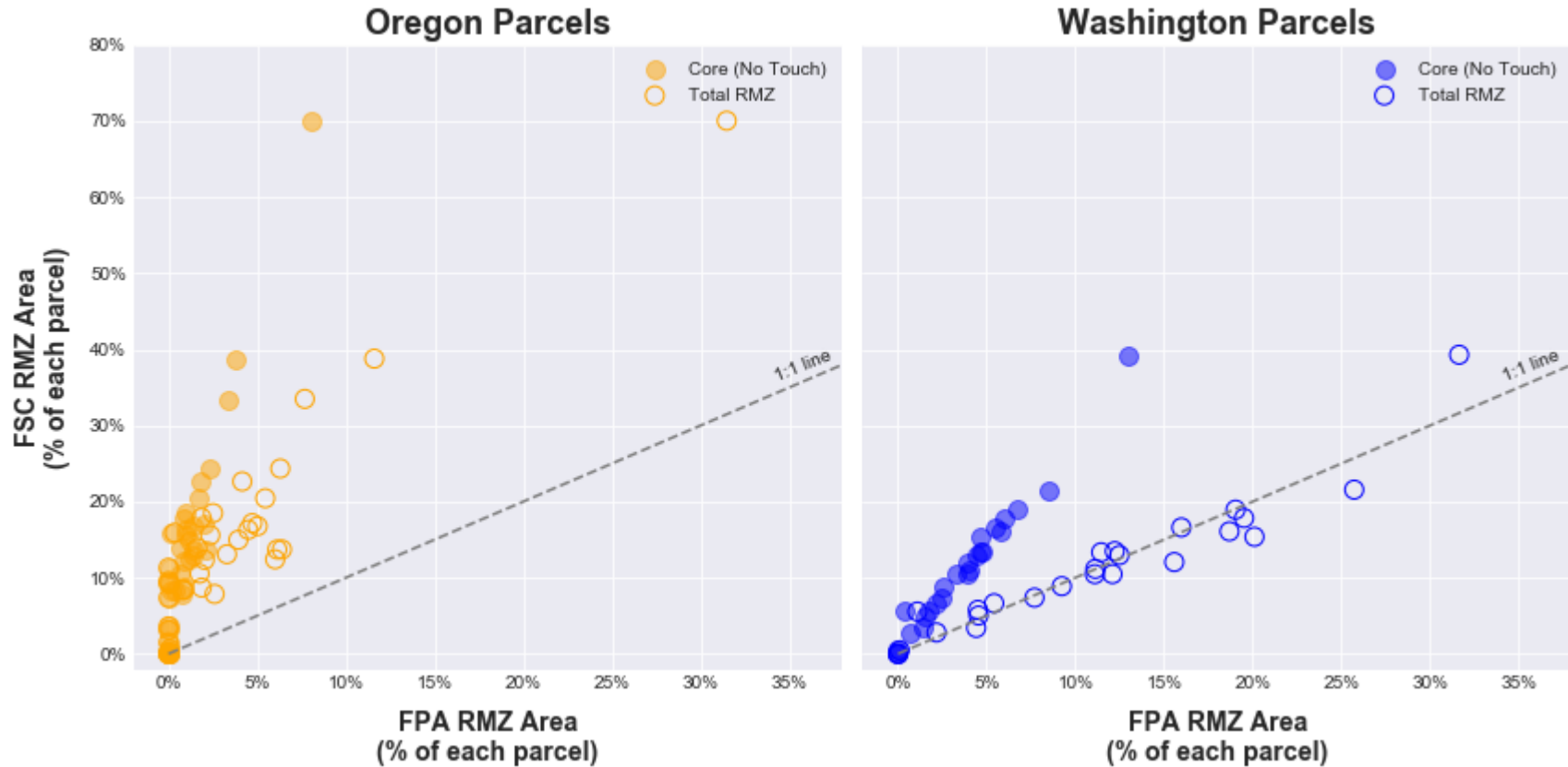


Figure 1.4. Area encumbered by Riparian Management Zones under FPA and FSC rules for Oregon and Washington.
These graphs show the percentage of each parcel that is covered under no-touch/core and total RMZs for both FSC and FPA rules. Points above/below the 1:1 line indicate RMZ buffers cover more/less area under FSC than FPA rules.

Carbon Storage in the Forest and Wood Products

To ensure relevance for policymakers considering new voluntary incentive or regulatory approaches to encourage forest carbon storage and sequestration, we quantify carbon storage in this study considering those carbon pools typically included in forest carbon accounting and offset crediting programs. These include above- and below-ground biomass of live trees, standing biomass of dead trees, and carbon storage retained in in-use harvested wood products. This approach is comparable to the accounting framework used by all major forest carbon accounting and crediting programs (e.g., California Air Resources Board (ARB) Forest Carbon Protocol, the Verified Carbon Standard, and the American Carbon Registry). Carbon offset accounting frameworks do not reflect a full Life Cycle Assessment approach and generally omit several carbon pools in the forest (e.g., downed dead wood) and following harvest removals (e.g., use of wood for energy, wood products remaining in landfills), as well as greenhouse gas emissions related to harvesting, transportation, wood products processing and distribution, or the combustion of harvest residues (slash burning). Nevertheless, we believe this carbon accounting framework corresponds to the most likely approach for quantifying parcel-level carbon reductions through forest carbon incentive policies under discussion in Oregon and Washington.

We calculate the carbon storage KPI as the average volume of CO₂-equivalent storage in the forest and harvested wood products pools, net of leakage due to market effects, over a 100-year timeframe. We generally follow the methods defined by the California ARB Forest Carbon Protocol (California Air Resources Board, 2015). We apply a single decay factor to account for the long-term average carbon storage in the harvested wood products pool. For this study region, this decay factor corresponds to 42.1% of the carbon removed from a stand in the merchantable portion of harvested trees being retained in wood products, on average, over 100 years. In addition, we follow the ARB Protocol to account for leakage due to market effects (referred to as “Secondary Effects Emissions” in the Protocol), which discounts any additional carbon stored in the forest if there is a decrease in timber output relative to BAU. This effect assumes that a portion of any decline in timber output from a parcel will be made up for by increased harvesting in other locations, resulting in a “leakage” of the additional carbon stored in the “project area”. The ARB Protocol assigns a 20% leakage factor to the difference of harvest removals in a given year between the projected scenario and a BAU (or “baseline”) scenario if the projected scenario has generated

a lower cumulative harvest volume than the baseline scenario up to that point in time. We diverge from the ARB Protocol's methods that establish a "baseline" for carbon offset crediting using a long-term average of the BAU scenario that must be at least as high as the average carbon storage in comparable forest types in an "Assessment Area". We do not impose a long-term average constraint on the BAU scenario for our calculation of additional carbon storage in alternative scenarios, but rather make these calculations using the dynamic values of carbon storage in both the BAU and the alternative management scenarios over the 100-year simulation timeframe.

Cumulative Timber Output

The timber output KPI is calculated as the cumulative sawlog volume (using Scribner boardfoot measure) of merchantable wood produced by each management scenario. Volumes are included for the following species: Douglas-fir, Sitka spruce (*Picea sitchensis* (Bong.) Carrière), western hemlock, grand fir (*Abies grandis* (Douglas ex D. Don) Lindl.), noble fir (*Abies procera* Rehder), Pacific silver fir (*Abies amabilis* (Douglas ex Loudon) Douglas ex Forbes), western redcedar, yellow cedar (*Callitropsis nootkatensis* (D. Don) Oerst. ex D.P. Little), red alder, and bigleaf maple (*Acer macrophyllum* Pursh).

Merchantable boardfoot volume is included for softwood trees with a minimum diameter at breast height (DBH) of 22.86 cm (9 in) to a minimum top diameter inside bark (DIB) of 15.24 cm (6 in). Merchantable boardfoot volumes are also included for two hardwood species—red alder, and bigleaf maple—with a minimum DBH of 27.94 cm (11 in) up to a minimum top DIB of 20.32 cm (8 in). Boardfoot volumes are determined using 9.75 m (32 ft) log equations for softwoods and 4.88 m (16 ft) log equations for hardwoods. All boardfoot volume calculations include a deduction for a 0.3 m (1 ft) stump. We also impose log volume adjustment factors to correct Scribner volume overestimation observed when using the National Volume Estimator Library (NVEL) Behre's hyperbola equations available in FVS (Wang, 2017). Boardfoot "defect" adjustment factors were calculated for several DBH classes for each merchantable tree species. The scales of these adjustments were determined by quantifying the average correction needed to have NVEL estimates match those produced by the regional volume equations utilized in the US Forest Service's Forest Inventory & Analysis Program (Waddell et al., 2014). These adjustment factors are presented in the Supplementary Material.

Discounted Cash Flow

For our third and final KPI, we calculate discounted cash flow using an annual discount rate of 5%, which was chosen based on recent industry presentations/reports (Blacklock, 2016; New Forests, 2013). Timber sale revenues are calculated using delivered log prices based on a recent log price report from the Washington Department of Natural Resources (2018) and are presented in Table 1.2. Management costs were determined based on a recent survey of industrial forest landowners in the Pacific Northwest (Arney, 2016) and are presented in Table 1.3. We calculate discounted cash flow using real discount rates and prices; inflation is not reflected in these calculations.

Table 1.2. Delivered log prices.

Species	\$/MBF
Douglas-fir	796
Sitka spruce	450
Western hemlock	640
Noble fir	640
Grand fir	640
Pacific silver fir	640
Yellow cedar	640
Western redcedar	1263
Red alder	852
Bigleaf maple	499

Notes: Derived from February 2018 log price report for Washington “Coast Marketing Area” (Washington Department of Natural Resources, 2018).

Table 1.3. Management costs.

Activity	\$	Per
General administration	86	ha/year
Site preparation	210	ha
Tree planting	0.73	seedling
Brush control (@ age 5)	334	ha
Harvest administration	5	MBF
Hauling	100	MBF
Road maintenance	15	MBF
<u>Ground-based harvest:</u>		
Regeneration harvest	150	MBF
Commercial thin	175	MBF
<u>Cable logging:</u>		
Regeneration harvest	200	MBF
Commercial thin	300	MBF

Embodied Carbon

We also present a hybrid indicator for the “embodied carbon” of wood products generated in each management scenario, calculated as the average carbon stored in the forest and wood products divided by the cumulative amount of timber produced over the modeling timeframe (100 years). This metric provides an indicator of the embedded carbon footprint for timber generated under each alternative management regime, which may be useful in the same context that greenhouse gas displacement factors can be used to quantify the benefits of utilizing wood in place

of more carbon-intensive building materials (Sathre and O'Connor, 2010), or in the context of Environmental Product Declarations, which seek to quantify the environmental impact of products commonly considered by builders.

Incentives or Price Premiums that Close the Financial Gap with BAU

Alternatives to BAU forest practices are generally expected to provide a lower return on investment given current market conditions and policies. In many forest carbon offset standards, this assumption is often explicitly identified as a financial barrier to the adoption of new practices and integrated into the definition and assessment of additionality (CDM - Executive Board, 2007; Verra, 2012; American Carbon Registry, 2018; von Hagen, 2009). A variety of alternative revenue strategies, including the sale of conservation easements and carbon offsets, as well as the delivery of higher-value logs from extended rotations or commercial thinning harvests timed to favorable market conditions are often considered as incremental approaches to help close and/or overcome these financial barriers (von Hagen, 2009).

We quantify two options for reducing the financial gap between BAU and alternative management scenarios. The first option we consider is a premium on wood sold. We quantify the premium on wood as a multiplier to the periodic gross revenue from the sale of timber. This could correspond, for example, to a premium based on the production of logs with a higher value per volume measure, or to a premium based on consumer willingness-to-pay for a third-party certification such as FSC. The second option we consider is an incentive for additional carbon stored in the forest and harvested wood products. In this approach, we add new revenue based on the difference in credited carbon stored (in the forest and wood products, net of leakage) between each alternative scenario and BAU. In the first five-year simulation period, the difference between an alternative management scenario and BAU is calculated and rewarded. In all successive periods, the change in carbon stored in each alternative management scenario versus the change of carbon stored in the BAU scenario is rewarded. The additional income provided by the wood premium or carbon value is discounted to present value. We utilized a simple optimization using the *scipy* Python package to search for the incentive value which minimizes the squared difference in NPV at the end of 100 years between an alternative management scenario and BAU for each parcel individually.

Growth-and-Yield Simulation

We conducted growth-and-yield modeling using the Pacific Northwest Coast (PN) variant of the Forest Vegetation Simulator (FVS) (Keyser, 2017; Dixon, 2017; Crookston and Dixon, 2005) with Database and ECON extensions employed to streamline data input/output and for economic calculations (Martin, 2013; Crookston et al., 2003).

We calibrated FVS default growth and mortality parameters based on comparison of growth projections of Douglas-fir plantations established from bare ground with historical yield tables (Chambers, 1980; Curtis et al., 1982; McArdle et al., 1961; Mitchell and Cameron, 1985; Schumacher, 1930; Stand Management Cooperative, 2016), long-term/permanent plot records (Curtis and Marshall, 2002; Williamson, 1963), and regional forest inventory data (USDA Forest Service Pacific Northwest Research Station, 2017) (Supplementary Material). We identified modifications for the maximum Stand Density Index (SDI_{max}) to adjust competition-related mortality rates, background mortality rates prior to age 30, and multipliers to reduce the annual basal area increment of large trees. These adjustments to default FVS-PN model behavior are our best attempt to strike a conservative balance capturing rapid growth and yield apparent from contemporary intensive Douglas-fir plantations (Stand Management Cooperative, 2016) for young stands without allowing the relative gains in growth and productivity in these young stands to persist into older age. This approach directs yields in stands older than those observed in intensive plantation field plots (Stand Management Cooperative, 2016) back towards the median range of values in Douglas-fir forests observed across the landscape of western Oregon and Washington from historical yield tables, permanent plots, and FIA data. By forcing the trajectory of advanced growth of younger plantations back down towards historical trends, we run the risk of underestimating the potential timber yield and harvest revenue these intensively managed stands might be able to sustain if they were allowed to grow to ages beyond “financial maturity”. This decision was made primarily because little to no inventory data are currently available for intensively-managed plantations using contemporary silvicultural practices and seed sources beyond the range of 40–50 years.

To identify the optimal rotation ages for each of our management scenarios (Table 1.), we simulated the establishment and growth of Douglas-fir plantations from bare ground on a relatively flat slope (10%) and a steep slope (40%) for each Douglas-fir 50-year Site Index value ranging from 15.2 m (50') to 48.7 m (160 ft) in increments of 1.5 m (5 ft). The age at which NPV peaked

for each slope-by-productivity combination was identified as the “financially optimal” rotation age used for our BAU and SHORT~FSC scenarios. Our calibration of FVS induced a distinct downward bend in timber volume growth around age 35–40 as stands approached SDI_{max} . We applied a moving average function to smooth the NPV curve derived from these simulations and identified financially optimal rotation ages of 38–44 years, depending on Site Index and slope (stands with steeper slopes had slightly later rotation ages than those on flat ground). The age of culmination for Mean Annual Increment (MAI) in terms of total cubic volume was identified as the “biologically optimal” rotation age for our LONG~FPA and LONG~FSC scenarios. These longer-rotation scenarios included commercial thinning to capture density-driven mortality after age 30 and resulted in a rotation age of 75 years for all stands regardless of Site Index.

We conducted a 100-year simulation using a 5-year time step for every management unit (out of 10,068) across the 64 parcels. We modeled every prescription available (4 non-riparian, 1 “grow only”, and 20 riparian prescriptions meeting the distinct retention requirements of each stream type, size, etc.) for each management unit, producing a total of 252,150 simulations. The outputs from these simulations were then scaled to the area assigned to each treatment for each stand.

Many of the KPIs and other indicators of interest across the parcels are not normally distributed. As such, we report the median and percentile ranges for these indicators rather than a mean and standard error.

Computing Environment for Simulations, Data Analysis, and Visualization

We ran FVS simulations on a 32-core Linux server using the latest version of open-fvs built from source (USDA Forest Service, Forest Management Service Center, 2015). We utilized PostgreSQL databases for FVS input and output to permit batch modeling with an asynchronous parallel processing workflow coded using the Python programming language. All code developed by the authors for implementing this processing workflow is open-source and has been published by Ecotrust on GitHub (Ecotrust, 2018). The parallel processing workflow enabled the completion of these 252,150 FVS simulations in 6–7 h. Data analysis and visualization has been conducted using packages in both the Python and R languages. The outputs of FVS simulations, expanded to a per-area basis for each parcel, are included in the GitHub repository, along with a Jupyter Notebook (Thomas et al., 2016) containing all the code necessary to generate the graphs and tables

included in this article. Utilizing the Binder service, which is linked from the landing page of the GitHub repository for this project (Ecotrust, 2018), a copy of this Jupyter Notebook can be launched in the cloud with the relevant computing requirements pre-installed and configured to allow for the reproduction of our quantitative analysis and visualizations without the need to download or install any software.

1.3 RESULTS

Calibration of FVS

Our calibration of the FVS-PN variant for Douglas-fir plantations against independent field data and yield tables revealed a consistent over-prediction of cubic and boardfoot volumes using default FVS parameterization. These over-predictions were observed at the individual tree level, as well as for volume growth and carrying capacity metrics at the stand level. The use of an uncalibrated model would have resulted in much greater estimates of future volume and carbon, particularly for older stands. The parameterization we identified through trial and error follows plantation growth curves fairly well through financial maturity (age 40), after which point the projected metrics gradually return into the range of average conditions observed in natural stands. Outputs from the default and calibrated FVS-PN projections are presented in the Supplementary Material for the height of dominant trees, number of trees per acre, basal area, quadratic mean diameter, cubic volume, Scribner boardfoot volume, Stand Density Index, and Curtis Relative Density. A series of yield tables for these metrics using the calibrated model are also provided.

Simulated Forest Dynamics

The optimal conversion time for most stands was identified as the first simulation cycle, with 77% of non-riparian acreage on Oregon parcels and 81% of non-riparian acreage in Washington converted in the first five years of simulation; over 90% of non-riparian areas in both states underwent the conversion harvest within 20 years of the start of the simulation. This rapid conversion resulted in saw-tooth outputs characteristic of stand-level simulations, with the age-based harvest triggers clearly apparent across all 64 parcels (Figure 1.5).

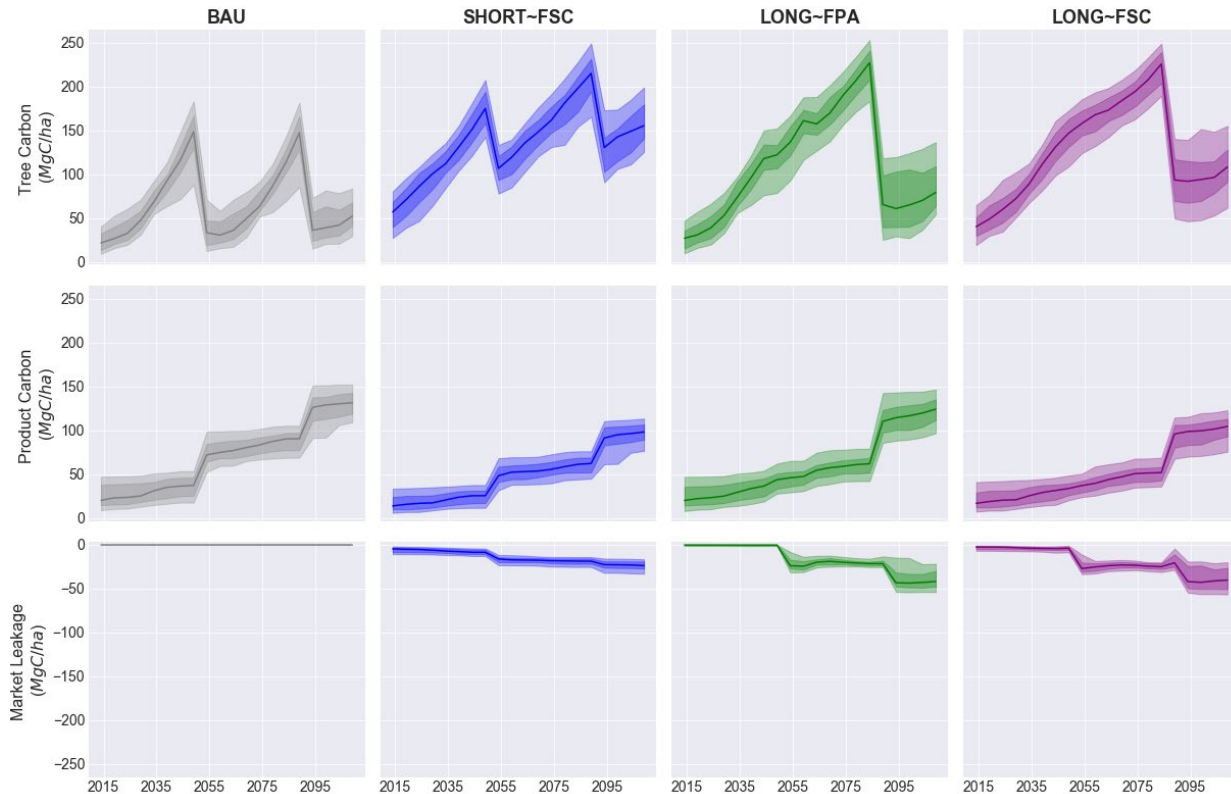


Figure 1.5. Carbon stored in three primary categories of pools/emissions under four alternative management scenarios.

The dark line in each graph represents the median value over time across the 64 parcels. The shaded areas correspond to the 25th–75th percentile range (darker), and the 10th–90th percentile range (lighter).

Over the 100-year simulation period, the greater green tree retention in ~FSC scenarios produced successive accumulation of carbon stored in trees over time, while the shorter ~40-year rotations of the BAU scenario generally reset the amount of carbon stored in trees to a low and consistent level at each harvest. The use of a fixed decay factor generated a step-wise increase for carbon stored in harvested wood products with each successive harvest for the ~40-year rotations (BAU and SHORT~FSC). For both the BAU and LONG~FPA scenarios, carbon stored in wood products began to rival that stored in trees by the end of the simulation; for both the ~FSC scenarios, the amount of carbon storage in trees was almost always higher than in products.

Leakage due to market effects was generally small compared to the amount of carbon stored in trees or products, and occasionally resulted in periodic calculations of modest positive market leakage. This positive leakage effect emerges in cases where an alternative management scenario has produced less cumulative timber output than the BAU scenario up to a certain point

in time, but generates a larger periodic output than BAU for one or more years, and can be seen for the LONG~FPA and LONG~FSC scenarios where cumulative market leakage follows brief upward movements (becomes less negative) for short periods (Figure 1.5).

An inspection of stand images produced using the Stand Visualization System (McGaughey, 2004) indicated that the relatively high levels of green tree retention in the SHORT~FSC scenario (30% at each regeneration harvest) often resulted in the development of two-aged stand structures by the end of 100-year simulation period, while the longer rotation and BAU scenarios usually maintained a single-aged stand structure (data not shown).

Key Performance Indicators

The dynamics of timber output, carbon storage, and discounted cash flow over the 100-year simulations for each scenario are displayed in Figure 1.6.

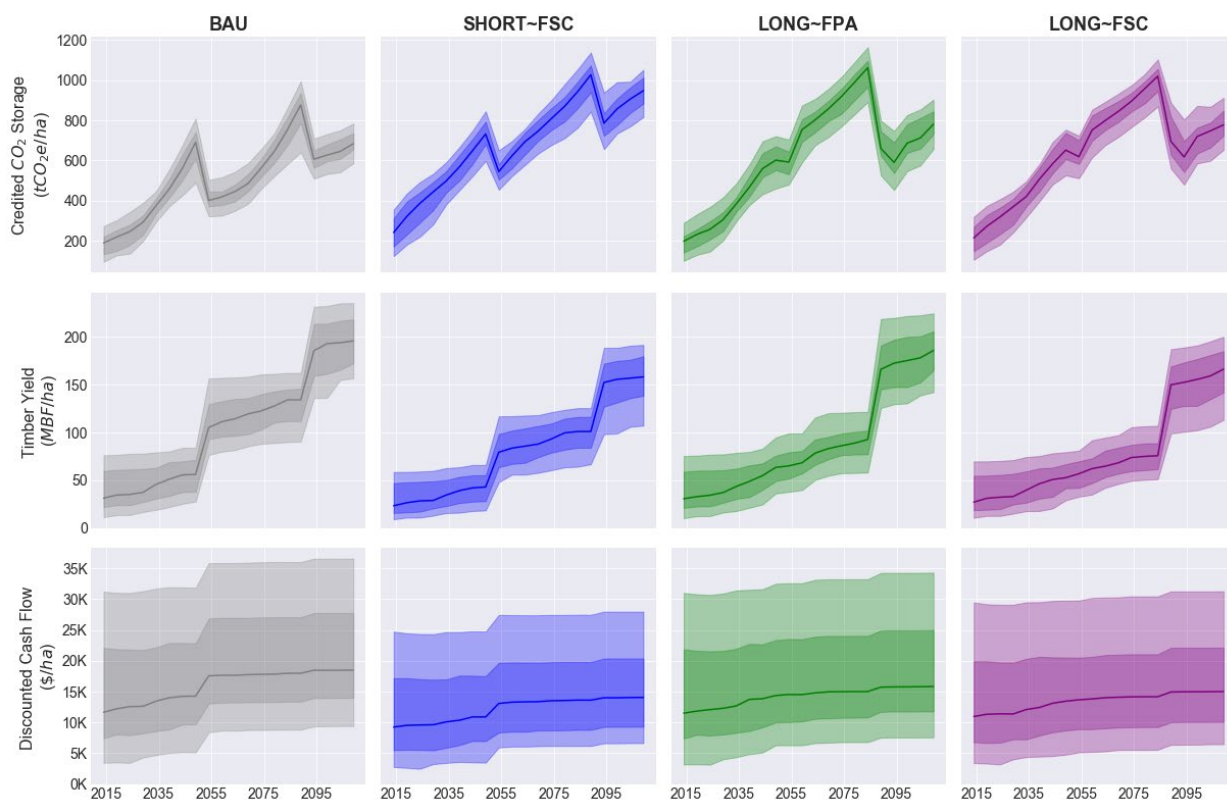


Figure 1.6. Dynamics of three KPIs over 100 years under four alternative management scenarios.

The dark line in each graph represents the median value over time across the 64 parcels. The shaded areas correspond to the 25th–75th percentile range (darker), and the 10th–90th percentile range (lighter).

In general, the BAU scenario usually performed the best under the two traditional forestry KPIs (timber output and discounted cash flow), and the worst under the carbon KPI. The distinct contributions to the KPIs by the individual constraints on green tree retention and riparian buffers can also be isolated and are displayed in Table 1.4. For example, a shift from ~FPA green tree retention levels in the BAU scenario to ~FSC retention levels showed a larger impact on carbon storage, timber output, and NPV than the expansion of riparian buffers did.

The long-rotation scenarios generated a lower yield of cumulative timber volume over time compared to the short-rotation scenarios. However, the long rotation scenarios carried a higher standing volume of harvestable timber than the shorter rotation scenarios. The relatively rapid conversion across the parcels and subsequent homogenization of age classes likely exaggerated an effect from the fact that the shorter rotations were able to get two full harvest cycles into the simulation timeframe while the long rotation scenarios did not. Considering the standing boardfoot volume in addition to the cumulative harvested volume by the end of the simulation period, the longer rotations showed higher totals: the median standing plus harvested timber volume was 288 MBF/ha among parcels for the LONG~FPA scenario and 268 MBF/ha under the BAU scenario; LONG~FSC had a median of 618 MBF/ha while SHORT~FSC had a median of 601 MBF/ha. These values demonstrate higher Mean Annual Increment for the long rotations compared to the short rotations. Furthermore, the different conclusions that might be drawn from considering yields alone versus Mean Annual Increment suggest that a more evenly distributed set of age classes across the landscape, often referred to as a “regulated” harvest schedule, would translate into higher yields for the longer-rotation scenarios. A more phased approach across the landscape to the conversion from initial forest conditions than the one we simulated may have demonstrated this case more directly.

The greater green tree retention and wider and more restrictive RMZ rules of the ~FSC scenarios corresponded with higher carbon storage at the expense of lower timber yields and NPV than their ~FPA counterparts within our 100-year timeframe. Compared to Oregon, Washington showed modestly higher carbon storage, lower yields, and lower NPV for all scenarios (based on median values of KPIs across the simulated parcels). Looking at the KPIs for the alternative scenarios relative to BAU across both States (Figure 1.7), LONG~FPA scenarios had median values of 22.5% more carbon stored, 4.3% less timber generated, and 11.7% lower NPV; LONG~FSC yielded 25.5% more carbon, 15.2% less timber, and a 20.6% lower NPV; and

SHORT~FSC yielded 32.4% more carbon, 19.0% less timber, and 25.3% lower NPV. LONG~FPA also provided a relatively tighter clustering of points compared to BAU than did either of the ~FSC scenarios. LONG~FPA had overall lower variation in stand-level response when compared to the ~FSC scenarios; the ~FSC scenarios increased the likelihood of generating higher carbon storage and lower NPV.

Per MBF of wood produced, ~FSC scenarios always had a larger embedded carbon value. In Oregon, BAU scenarios generated carbon storage of 2.4 tCO₂e per MBF produced over 100 years (median of the 64 simulated parcels), while SHORT~FSC, LONG~FSC, and LONG~FPA stored 4.2, 3.9, and 3.1 tCO₂e per MBF. In Washington, BAU stored 2.9 tCO₂e per MBF, while SHORT~FSC, LONG~FSC, and LONG~FPA stored 4.1, 3.9, and 3.7 tCO₂e per MBF. This indicates mid-range estimates for an implicit embedded carbon benefit of 1.5 to 1.8 (Oregon) or 1.0 to 2.1 (Washington) of tCO₂e per MBF for logs produced under ~FSC constraints relative to those produced under current common practice.

Values that Close the Financial Gap with BAU

The values of a prospective premium on wood products or carbon incentive program required to close the financial gap between the alternative management scenarios and the BAU scenario are shown in Table 1.5. In general, LONG~FPA would be able to break-even with BAU with the support of more modest premiums or incentives than those required for either of the ~FSC scenarios.

Table 1.4. Isolated impacts of green tree retention and RMZ widths constraints on KPIs for each scenario.

	SHORT~FPA			SHORT~FSC			LONG~FPA			LONG~FSC		
	<i>MBF</i>	<i>tCO_{2e}</i>	<i>\$K</i>	<i>MBF</i>	<i>tCO_{2e}</i>	<i>\$K</i>	<i>MBF</i>	<i>tCO_{2e}</i>	<i>\$K</i>	<i>MBF</i>	<i>tCO_{2e}</i>	<i>\$K</i>
OREGON												
~FPA Buffers	199	497	19.1	169	634	15.3	197	608	17.1	183	595	15.9
~FSC Buffers	176	553	18.0	153	679	14.3	163	686	15.7	168	646	15.3
WASHINGTON												
~FPA Buffers	185	518	18.0	166	663	14.6	174	639	15.6	170	620	15.5
~FSC Buffers	178	527	17.2	159	659	14.0	158	656	13.9	165	616	14.4

Each set of six cells displays the median of each KPI on a per hectare basis over the 100-year simulation period for each management scenario, with riparian rules following either ~FPA or ~FSC. \$K refers to thousands of US dollars of Net Present Value. Both ~FPA scenarios involve the same green tree retention (10 tph/4 tpa) during regeneration harvests, while SHORT~FSC retains 30% of pre-harvest basal area and LONG~FSC retains 10%. To quantify the isolated effects of imposing 30% green tree retention versus minimum ~FPA retention for a given state, for example, compare the first row beneath SHORT~FPA with those in the first row beneath SHORT~FSC.

Table 1.5. Additional value for wood and carbon storage required to close the financial gap of alternative management scenarios with the BAU scenario.

	SHORT~FSC			LONG~FSC			LONG~FPA		
	<i>25%</i>	<i>median</i>	<i>75%</i>	<i>25%</i>	<i>median</i>	<i>75%</i>	<i>25%</i>	<i>median</i>	<i>75%</i>
OREGON									
Wood Premium (%)	9.8	15.0	21.3	3.0	5.2	9.8	0.0	1.4	2.5
Carbon Value (\$/tCO _{2e})	27.90	41.21	49.43	30.84	42.68	51.50	0.00	17.94	23.99
WASHINGTON									
Wood Premium (%)	8.7	10.7	12.0	5.1	5.9	9.6	2.3	4.7	6.2
Carbon Value (\$/tCO _{2e})	10.96	26.03	39.56	26.91	32.76	40.03	20.61	33.83	38.02

Values shown are the 25th, 50th, and 75th percentiles for parcels in each state.

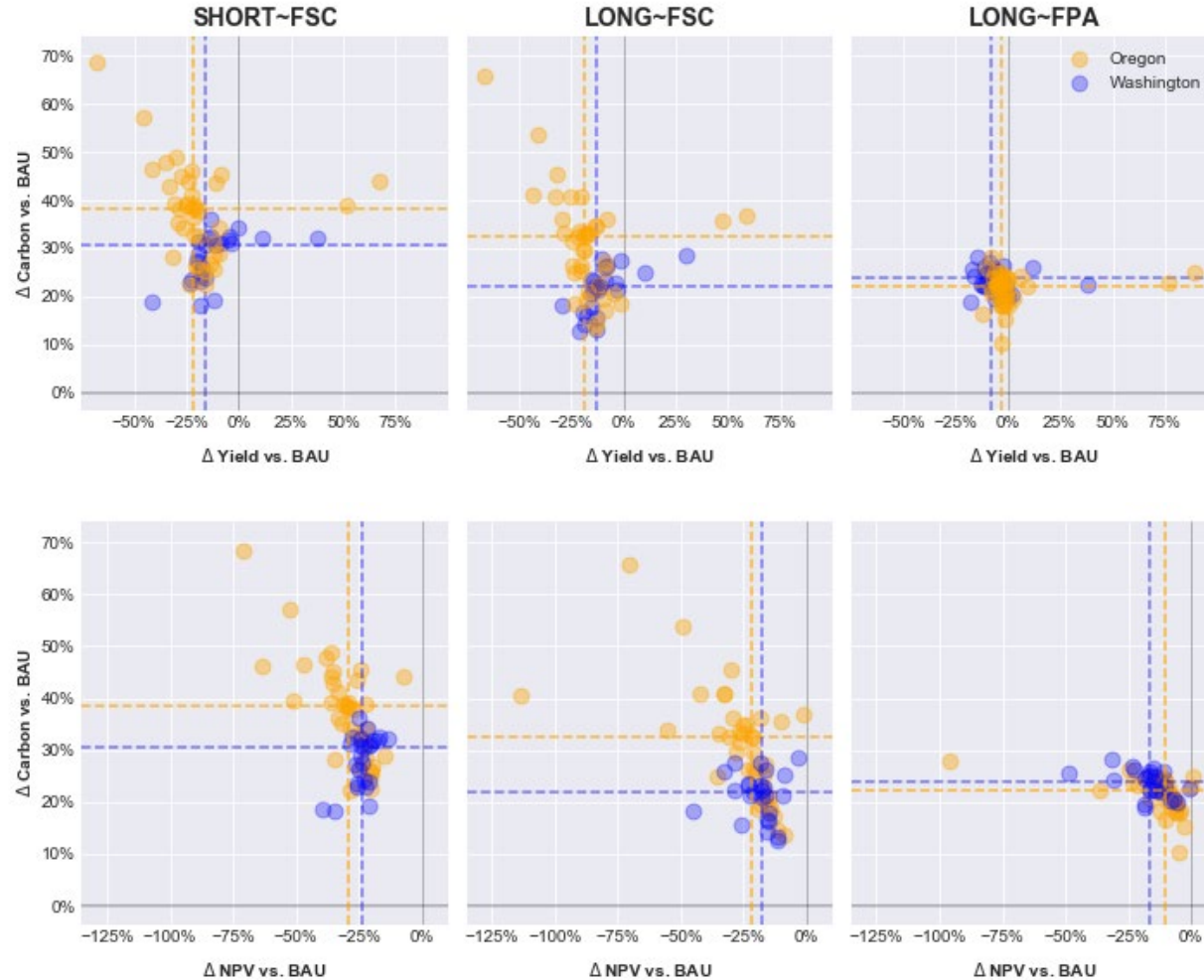


Figure 1.7. Shifts in timber yield, carbon storage, and cash flow relative to business-as-usual (NPV~FPA) for each of the three alternative scenarios.

The top row of graphs shows carbon and timber yield shifts; the second row shows carbon and cash flow shifts. Each point corresponds to one of the 64 simulated parcels. The dashed lines represent the median value

1.4 DISCUSSION

Fundamental Importance of FVS Calibration

Under default parameterizations, the FVS-PN variant projected wildly unrealistic volume and value growth for the Douglas-fir plantation scenarios considered in this study. Comparisons of projected volumes with independent datasets including historical yield tables, extensive publicly-available forest inventory data, and long-term permanent plot records allowed the growth-and-yield projections to be reined in significantly. This required the authors to make explicit decisions and balance tradeoffs in model performance to reasonably capture growth in young and mature Douglas-fir management scenarios of interest. Although the developers of FVS have encouraged and suggested quality control and quality assessment best practices to calibrate or validate the model (Vandendriesche, 2010), and a host of calibration and validation examples exist in the literature, the potential scale of errors from the use of FVS default settings are not common knowledge among model users. Studies intended to be policy- or management-relevant using FVS without describing how they calibrated or validated the model may be vulnerable to large errors, a vulnerability that may be particularly concerning if climate change factors are integrated into FVS projections where little to no independent validation data may be available. For example, although forest carbon offset programs often impose substantial costs to minimize the potential variability or error in field measurement, they tend to ignore much larger sources of variability associated with the use of growth-and-yield models over a century-scale timeframe, whether or not climate change is acknowledged (Fischer et al., 2017).

The calibration effort we present in the Supplementary Materials strikes a careful balance between estimates of rapid plantation growth at younger ages returning to landscape-level averages over time. These tradeoffs were considered explicitly with the goal of increasing our confidence that the values we were primarily interested in quantifying (e.g., the tradeoffs in carbon storage and timber output between long- and short-rotation scenarios) could be done conservatively, that is, with a reduced likelihood for overestimation of the effects of alternative management.

Tradeoffs in Timber Production, Carbon Storage, and Cash Flow

The observation that all alternative scenarios, for every parcel evaluated, stored a substantially higher volume of carbon in the forest and in harvested wood products highlights the

potential for a variety of management options to improve the carbon storage performance of Douglas-fir management beyond BAU practices. However, the financial analysis we present highlights the significant financial hurdle that these alternative management approaches face. Our findings of NPV under ~FSC scenarios on the order of 25% lower than BAU is consistent, but on the low end of estimates produced by others who have considered a broader suite of constraints required for FSC certification for Pacific Northwest Douglas-fir management (Tóth et al., 2013; Mendell and Lang, 2013).

The tighter clustering of LONG~FPA outcomes relative to the ~FSC scenarios (Figure 1.7) indicates that the greater retention of trees both in- and out-side of RMZs under the ~FSC scenarios (including species other than Douglas-fir), may introduce a substantial amount of variability in the projected outcomes for all KPIs. The diversification of species and size classes may offer both positive and negative aspects to future management outcomes. Any impacts of natural disturbances or climatic change, which are not simulated here, would be expected to produce different effects in homogenous and heterogenous stands; the more tightly predictable outcomes for a homogenous stand may also involve a greater exposure to uncertain future risks.

Multiple Financial Gap-Closing Strategies Would Likely Need to be Used for FSC to Compete with BAU

The analysis of options to close the financial gap between alternative management scenarios and BAU offers a practical set of indicators that inform their financial desirability (and presumably, the likelihood and scale with which these alternative practices could be implemented with supportive market conditions or policies). A premium on wood produced under these alternative management scenarios on the order of 15%–20% could be generated from the sale of higher quality larger logs produced by longer rotations as well as consumer willingness to pay for third-party certification, although neither of these options are consistently available in the market and both are likely to involve substantial marketing challenges.

Production and price premiums for higher-quality appearance-grade wood from larger-diameter logs have declined in the Pacific Northwest since the 1960s with the transition away from harvesting old-growth and toward shorter rotations of plantations (Haynes, 2003, 2005). Premiums for larger-sized and higher-quality logs now occur sporadically and fluctuate significantly with both domestic and international market conditions. In the absence of sustained demand, larger logs

now commonly face a price penalty rather than a premium, as many sawmills in the region have evolved to more efficiently process the supply of smaller logs more typical from the region's industrial forestlands. Premiums for export-quality or other high-value domestic log sorts are thus likely to offer landowners relatively unpredictable and short-lived opportunities to close the financial gap of alternatives to BAU management over time.

A growing interest among green builders for sustainable wood products may also offer a complementary value stream, although the potential demand for certification among green builders remains poorly understood. There is generally little evidence from the Pacific Northwest of price premiums for FSC-certified wood sometimes present in the retail setting translating up the supply chain into higher log prices received by FSC-certified landowners.

It is important to note that the options to close the financial gap between alternative management scenarios and BAU represent the value delivered to the landowner, not the retail price or premium paid by a consumer. This implies that any additional production or transaction costs required to deliver that value to the landowner would require even higher prices and premiums than those estimated here. For example, the market price of carbon required to close the financial gap between ~FSC scenarios and the BAU scenario would need to reflect the high startup and transaction costs for carbon market participation that present a substantial financial barrier to broader carbon market participation, particularly for smaller landholders (Kerchner and Keeton, 2015; Galik et al., 2012; Cacho et al., 2013, 2004). For example, an analysis of the break-even carbon price for a 65-year rotation to match the NPV of a 45-year rotation in similar forest conditions that also included a more thorough accounting of carbon project development costs was \$49.87 per tCO₂e (Fischer et al., 2017). These carbon prices are several times higher than those observed in prevailing markets (Hamrick and Gallant, 2017), and highlight what van Kooten observed more than twenty years ago: "In general, inclusion of the external benefits from carbon uptake results in rotation ages only a bit longer than the financial (Faustmann) rotation age" (van Kooten et al., 1995).

State FPA Rules Fundamentally Affect the Landscape and New Policy Opportunities

The difference between current State FPA regulations, particularly regarding riparian protections, have important consequences for financial performance, carbon storage, and the potential for new market- and non-market-based forestry policies to influence how the land is managed. The gap between FPA and FSC rules in Washington is narrower than it is in Oregon, which was clearly shown by the KPIs among our ~FSC and ~FPA simulations. The permission of more intensive harvest practices under Oregon FPA rules creates a higher opportunity cost for alternative management scenarios than is the case in Washington. As such, higher premiums for wood and higher values for carbon storage would be needed to shift forest management practices in Oregon compared to Washington in the absence of any changes to forest practices regulations.

In general, policies encouraging or incentivizing increased riparian protections, green tree retention, or the extension of rotation ages are likely to translate into greater carbon storage. A comparison of the isolated effects of the RMZ extension and increased green tree retention constraints demonstrates the expanded riparian protections offered a smaller impact on timber output and financial performance than would be accomplished by increasing FPA green tree retention levels from current minimum requirements to the 10 or 30% of pre-harvest basal area imposed by FSC. However, the greater decrease in timber output and financial performance involved with increased green tree retention also directly translated into proportionally higher carbon storage. Both green tree retention and greater RMZ protections are likely to correspond to other additional values including water quality and habitat for fish and wildlife that are not quantified here. Although carbon storage is the easiest of these ecosystem services to quantify, it is often these other values which form the basis for policies encouraging conservation practices. Additional carbon storage could be more reasonably considered as a co-benefit rather than the primary objective or indicator for public investments or regulation to encourage these practices.

FSC-Certification Appears to Offer a Clear Surrogate for Increased Forest Carbon Storage

The greater green tree retention and RMZ protections involved in FSC certification translated directly into higher average carbon storage in every parcel we evaluated, even after

accounting for market effects of leakage. The approaches we have taken to modeling the policies and growth dynamics for these managed forests indicate that FSC-certified wood carries an embedded carbon benefit relative to wood produced by current common practice in the region. Our analysis does not employ a full Life Cycle Analysis, as is commonly involved in Environmental Product Declarations, but nevertheless offers a clear indicator that FSC-certification offers a reasonable assurance of improved carbon balance compared to BAU.

Considering the high transaction costs that have formed a substantial barrier to participation in carbon markets (Fischer et al., 2017; Bouslah et al., 2010; Kerchner and Keeton, 2015; Galik et al., 2012; Cacho et al., 2013; Charnley et al., 2010), FSC-certification may offer a simple and cost-effective surrogate for directly identifying and rewarding landowners who follow forest practices that sequester additional carbon relative to BAU (and that also are likely to generate additional ecosystem service values, such as improved water quality and wildlife habitat). FSC-certification has been much more accessible to non-industrial landowners than carbon markets have been to date, and simpler incentive models similar to existing US Farm Bill programs may offer a more accessible approach to engage more landowners in forest conservation and management. Conservation incentive and carbon rental approaches have been considered before in the literature (Cacho et al., 2013, 2004; Bigsby, 2009), proposed recently by collaborative forest policy working groups (Forest Climate Working Group, 2015; Pinchot Institute for Conservation, 2011) and may be particularly relevant for new policy development if increasing accessibility to landowners or increasing the area of forestland engaged in forest carbon programs are important policy objectives.

Lower Variability in LONG~FPA Suggest Extending Rotation Ages Alone as a Viable Option

The LONG~FPA scenarios showed reduced variability in timber, carbon, and financial KPIs than either of the ~FSC scenarios. We believe this reduced variability is likely to correspond to the lower green tree retention, particularly for species other than Douglas-fir, in the LONG~FPA scenario compared to the ~FSC scenarios. The timber premium and carbon values required to close the financial gap with BAU were much smaller for LONG~FPA, and a substantial number of parcels, particularly in Oregon, were projected to perform better than BAU under all three KPIs

(as demonstrated by a timber premium of \$0 for 25% of Oregon parcels to meet or exceed the NPV of the BAU scenario).

Our simulations suggest that landowners pursuing a LONG~FPA scenario may have less uncertainty in the outcomes for these KPIs than those pursuing the SHORT~FSC or LONG~FSC scenarios. Drawing more general conclusions about the long-term effects of management alternatives on forest risks and vulnerabilities would need to extend beyond those values considered here and be difficult to quantify. The homogenization of existing forest composition may involve increased exposure to risks from maintaining a single-aged monoculture of Douglas-fir such as pests or diseases, or regeneration failure due to severe weather or drought. Longer rotations carry a larger timber value over an extended period and may thus entail additional risks for potential damage from fire, insects, disease, or severe weather, all of which may be affected by climate change (Agne et al., 2018). Management actions that diversify forest structure and composition may also alter a forest's resistance and resilience to environmental changes and disturbances. Ultimately, landowners, forest managers, and policymakers will continue to navigate and judge complex interactions among these risks, value systems, and the uncertainty of our changing climate into the future. The research presented here offers insights into the tradeoffs and options for balancing a few of the Key Performance Indicators now commonly being considered in management and policy decisions.

1.5 CONCLUSIONS

Our work clearly demonstrates that the adoption of certain forest practices including expanded riparian protections, increased green tree retention, and the extension of rotation ages can translate into substantially higher carbon storage than contemporary common practice for Douglas-fir management in the Pacific Northwest. However, these practices, particularly if adopted in combination, will also generally correspond to reductions in financial viability that may be important considerations for landowners, forest managers, and policy makers to quantify and balance. Estimates for the balance between these tradeoffs may be particularly sensitive to choices in growth-and-yield or other modeling approaches employed, and clear documentation of the assumptions and calibration of models used for these purposes should be considered a critical component for science- or evidence-based policy or forest management decisions.

The combination of forest practices required for FSC certification always stored more carbon than business-as-usual and produced an embodied carbon benefit that may offer a useful indicator for wood consumers concerned with quantifying and reducing the impact of their purchasing decisions, independent of any potential benefits derived from the substitution of wood for more carbon- or energy-intensive building materials. FSC certification also offers a useful case illustrating the potential for certification options simpler than current carbon offset crediting programs for market- and non-market-based policies to identify and incentivize landowners for additional forest carbon sequestration.

1.6 SUPPLEMENTARY MATERIALS

Modifications to FVS growth-and-yield: Cross-validation with yield curves, permanent plots, and FIA data

Ten simulations were conducted for each level of 50-year Site Index ranging from 22.86-39.62 m (75-130 ft) in 1.52 m (5 ft) intervals (i.e., 75, 80, 85... 130 ft), covering Site Classes II - IV. Simulations used default settings for the FVS Pacific Northwest Coast (PN) Variant (Keyser, 2017), and were iteratively adjusted through trial-and-error, resulting in the modifications below. Stochastic variability was introduced into the FVS model using the RANNSEED keyword. Stands were established with 1,075 trees per hectare (435 trees per acre (TPA)) of Douglas-fir with 85% survival rate.

Growth and mortality adjustments

Basal area increment was adjusted to a fraction of default rates using the BAIMULT keyword, and Maximum Stand Density Index, which governs density-driven mortality in FVS-PN was set at:

Table S1.6. FVS keyword modifications

Site Class	IV	III	II
SDIMAX	450	450	550
BAIMULT	0.3	0.5	0.5

For all stands, background mortality of large trees was modified to 5% of default mortality rates until the age of 30 using the MORTMULT keyword. At age 30, background mortality rates are restored to 100% of FVS-PN defaults. MORTMULT modifications do not affect density-related mortality, which is governed by the SDIMAX keywords described above.

Timber volume adjustments

By default, FVS calculates cubic and boardfoot volumes using equations from the National Volume Estimator Library (NVEL), which is documented by Wang (2017). The NVEL boardfoot volume were modified using the BFDEFECT keyword in FVS to provide timber volume estimates more consistent with field experiences and inventory census such as the US Forest Service Forest

Inventory Analysis (FIA), which place a greater emphasis on quantifying merchantable timber volume using regional equations. Adjustment factors to NVEL volumes were determined by computing Scribner boardfoot volume for all trees ≥ 15 " DBH of common merchantable species from FIA plots in western Oregon and western Washington using two separate methods: the NVEL equations, and the equations utilized by the FIA program (Waddell et al., 2014). In general, NVEL overestimated Scribner boardfoot volume compared to FIA equations, ranging from 5% in 15-20" red alder trees up to 49% in western redcedar. The correction factors used are shown in Table S1.6.

Based on comparisons to yields of Douglas-fir using the Stand Management Cooperative's (SMC) Yield Calculator (Stand Management Cooperative, 2016), we capped the Scribner volume adjustment factor at 25% for 25.4 cm (10 in) DBH trees. This allowed for standing boardfoot volumes to follow rapid plantation development consistent with SMC reference plantation network, while allowing volume corrections in larger trees that bring the yield curves down into the range of the 75th percentile of FIA observations as the stands reach 60-80 years old, and near the median value of FIA observations around 100 years old. This approach to boardfoot volume adjustments implies that the advanced volume growth observed in young plantations does not persist beyond the ages at which these stands are commonly harvested in contemporary production forestry in western Oregon and Washington. This is intended to be a conservative assumption and may underestimate the boardfoot volume that might be achieved if contemporary plantations were retained through longer rotations.

General notes about the following graphs and yield tables

In all graphs shown below, the FVS simulations are displayed as gray lines. Blue lines show data from permanent plots in fully-stocked Douglas-fir stands (Curtis and Marshall, 2002; Williamson, 1963). Black lines show published yield curves of Douglas-fir from California to British Columbia (Chambers, 1980; Curtis et al., 1982; McArdle et al., 1961; Mitchell and Cameron, 1985; Schumacher, 1930; Stand Management Cooperative, 2016). The red line in each graph represents the median value observed among "conditions" observed in the US Forest Service's Forest Inventory and Analysis (FIA) network of plots from western Oregon and Washington that contained more than 80% of their basal area in Douglas-fir. Site Indices recorded in the FIA data were translated directly into Site Classes for each "condition" (rather than using the cubic volume productivity classes employed by USFS). The red shaded areas around the FIA

Table S1.7. Mean percent defect incorporated into FVS for each species diameter at breast height

Species	25.4 cm (10")		38.1 cm (15")		50.8 cm (20")		63.5cm (25")	
	Defect	Count	Defect	Count	Defect	Count	Defect	Count
<i>Pseudotsuga menziesii</i> (Douglas-fir)	50%	23,575	33%	10,838	28%	5,443	33%	31,043
<i>Tsuga heterophylla</i> (western hemlock)	42%	9,554	22%	4,135	18%	1,905	28%	7,075
<i>Abies amabilis</i> (Pacific silver fir)	44%	4,241	22%	1,778	16%	861	15%	2,900
<i>Alnus rubra</i> (red alder)	37%	3,497	5%	1,169	11%	264	22%	112
<i>Thuja plicata</i> (western redcedar)	56%	1,673	45%	912	44%	489	49%	3,413
<i>Abies procera</i> (noble fir)	47%	706	24%	334	18%	192	14%	765
<i>Abies grandis</i> (grand fir)	39%	562	18%	271	14%	131	14%	255
<i>Picea sitchensis</i> (Sitka spruce)	54%	324	37%	182	34%	107	31%	600
<i>Callitropsis nootkatensis</i> (yellow cedar)	57%	226	46%	93	48%	45	52%	203

median line correspond to observations in the 25th – 75th percentile range (darker red) and the 10th – 90th percentile range (lighter red). Data from the PNW FIA Database were analyzed using a site index base age of 50 (assuming the majority of base ages actually recorded in the database listing a base age of 100 were incorrect (personal communication, Andy Gray, Washington state coordinator for the FIA program).

Boardfoot volumes are Scribner 32-foot log scale. Each data source uses different specifications for minimum top diameter inside bark. FVS, FIA, and SMC volumes are up to a 15.24 cm (6") minimum top diameter with a 0.30 m (1') stump. FVS, FIA, and SMC cubic volumes are shown for trees to a 10.16 cm (4") minimum top diameter inside bark with a 0.30 m (1') stump, while cubic volumes from other yield table publications are showing total cubic volume including top and stump.

The Yield Tables are direct outputs from the simulations conducted using the modified version of FVS-PN. The values in each cell represent the mean value for the metric being among the 10 simulations of each Site Index.

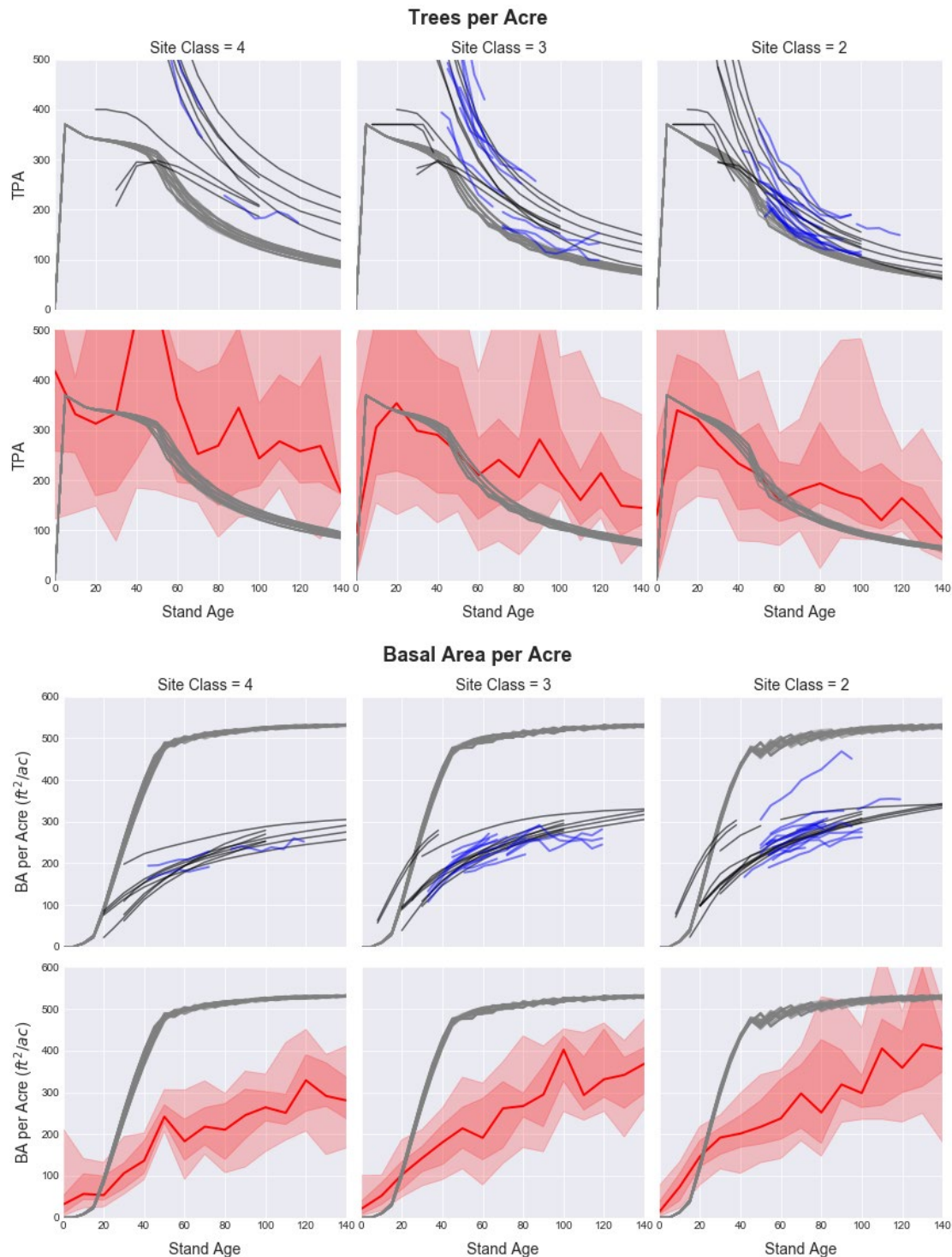


Figure S1.8. Default FVS Parameterization comparing PN variant output to measured stands (Part 1 of 3).

Trees Per Acre (TPA, top 6 graphs) and Basal Area (BA) per Acre (bottom 6 graphs). Gray lines show FVS simulations. Blue lines show permanent plot data. Black lines show published yield curves. Red line shows median, and bands show 25th-75th and 10th-90th percentile range of FIA data.

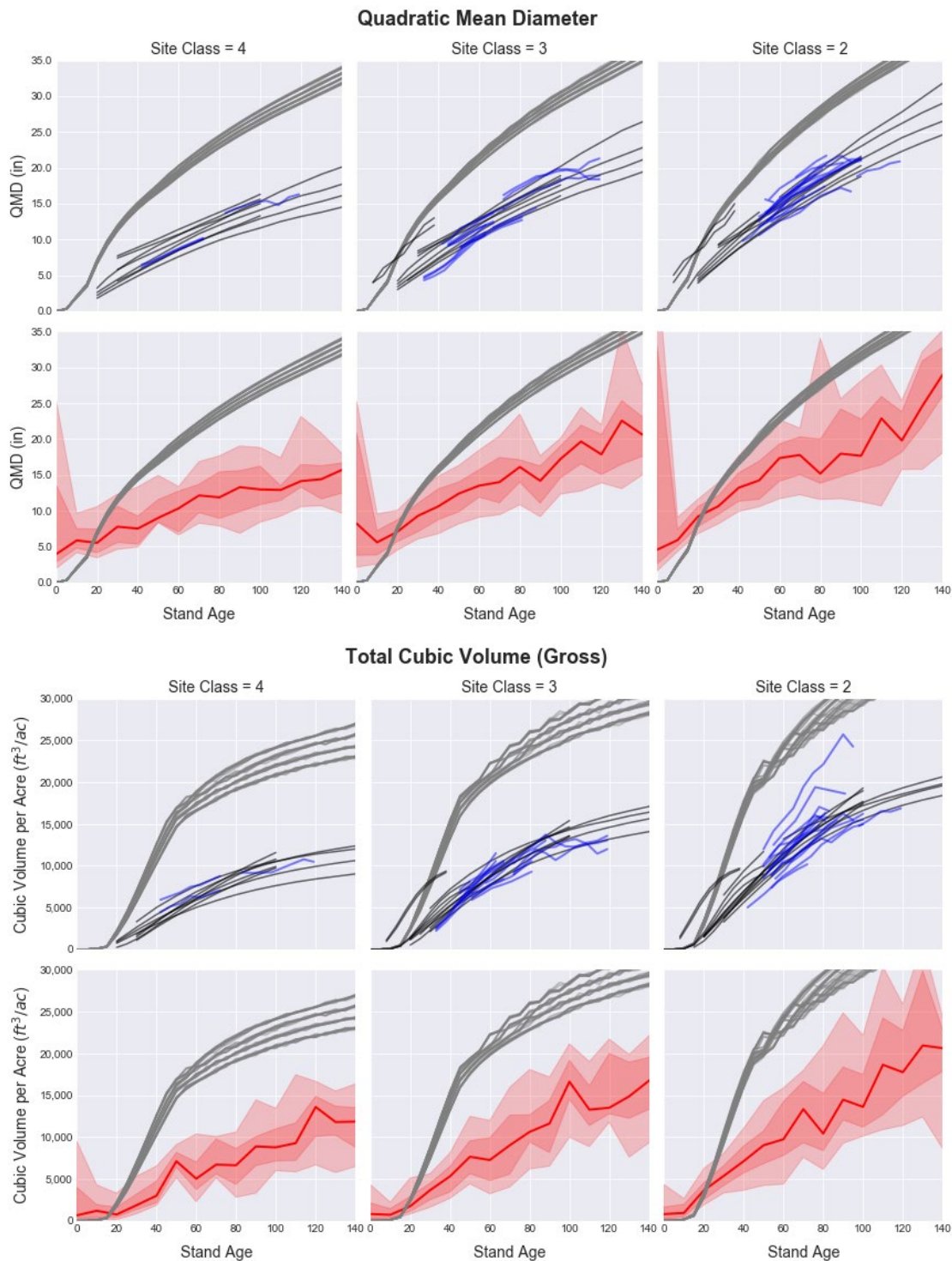


Figure S1.9. Default FVS Parameterization (Part 2 of 3). *Quadratic Mean Diameter (QMD, top 6 graphs) and Total Cubic Volume (bottom 6 graphs). Gray lines show FVS simulations. Blue lines show permanent plot data. Black lines show published yield curves. Red line shows median, and bands show 25th-75th and 10th-90th percentile range of FIA data.*

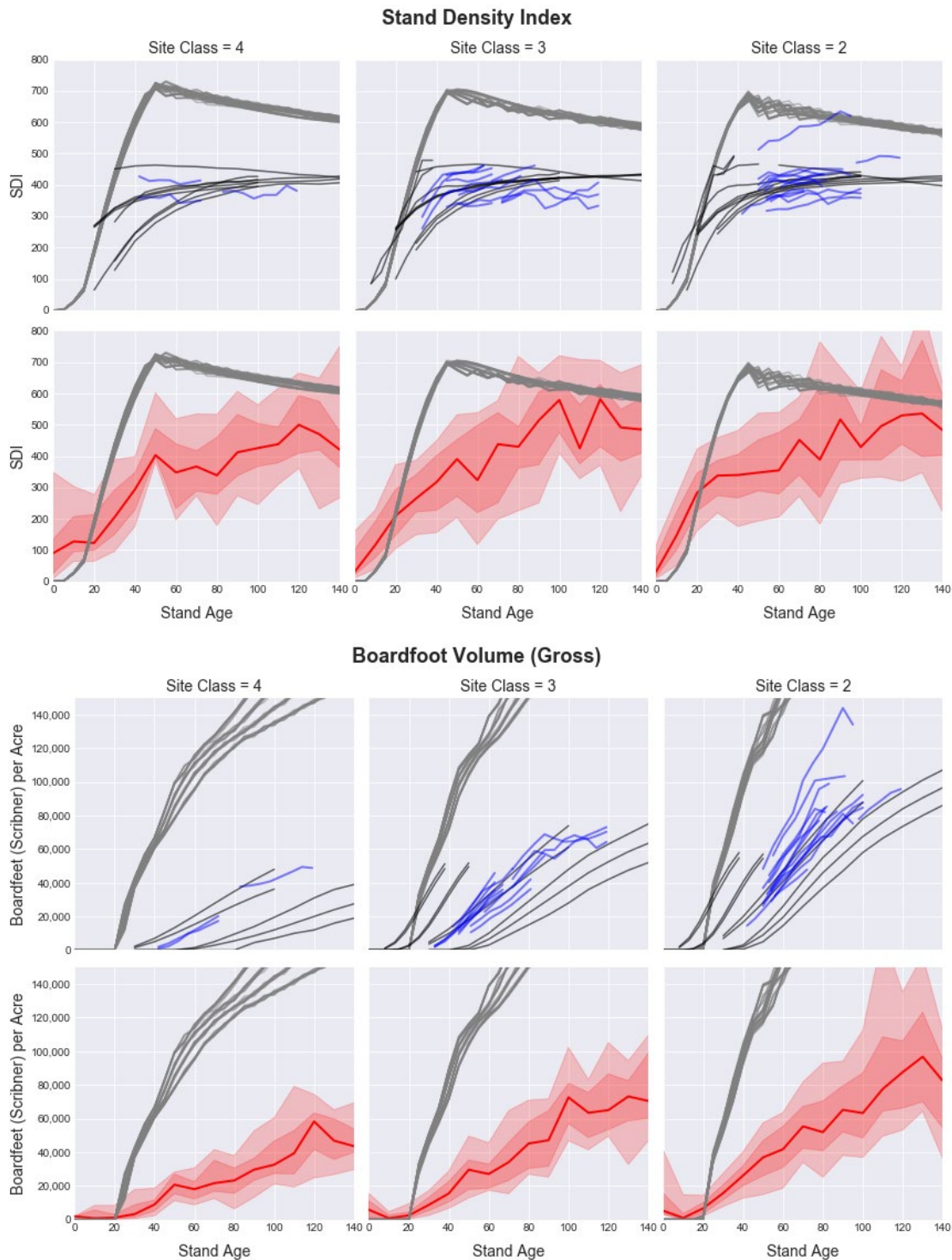


Figure S1.10. Default FVS Parameterization (Part 3 of 3).

Stand Density Index (SDI, top 6 graphs) and Gross Boardfoot Volume (bottom 6 graphs). Gray lines show FVS simulations. Blue lines show permanent plot data. Black lines show published yield curves. Red line shows median, and bands show 25th-75th and 10th-90th percentile range of FIA data.

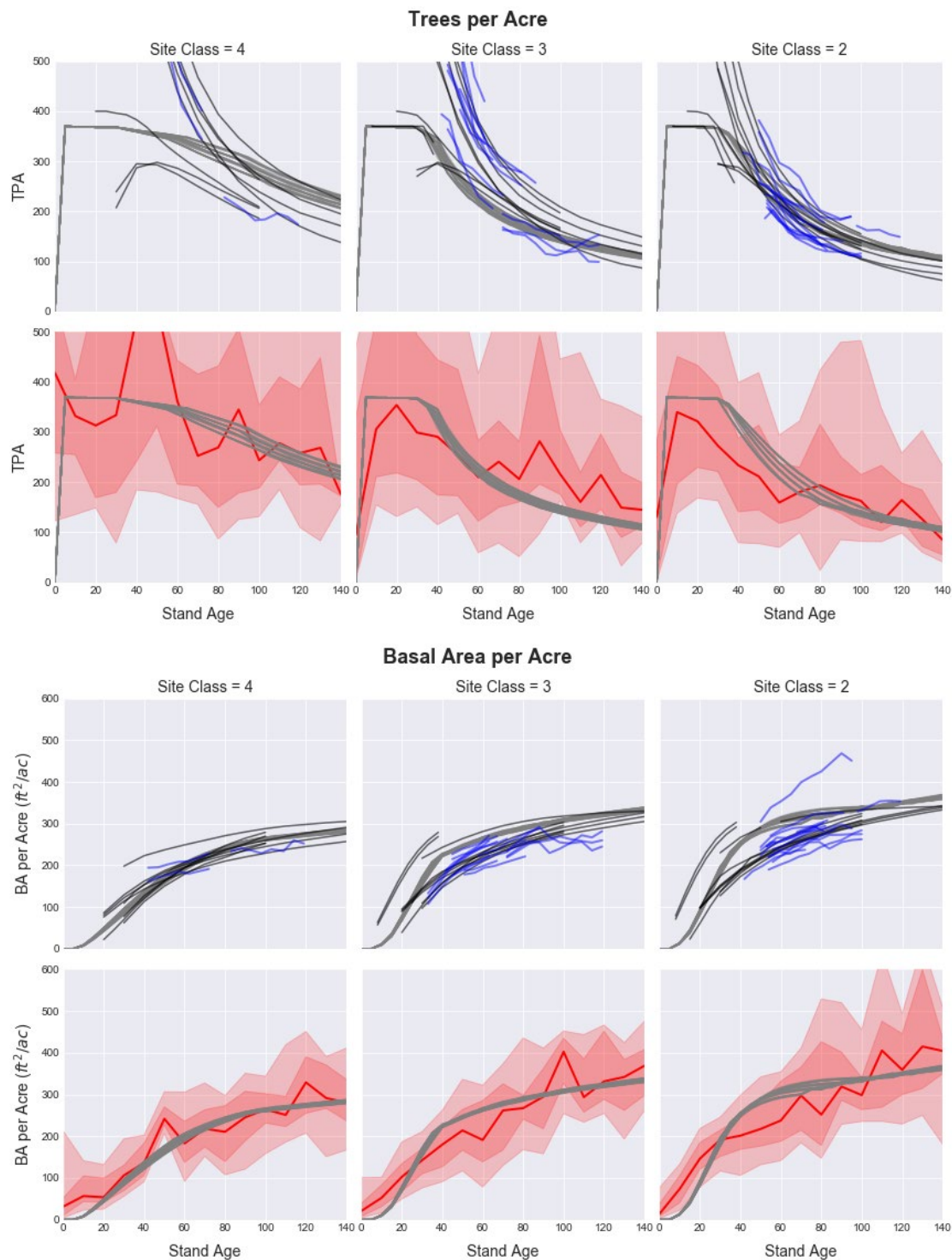


Figure S1.11. Modified FVS Parameterization (Part 1 of 3).

Trees Per Acre (TPA, top 6 graphs) and Basal Area (BA) per Acre (bottom 6 graphs). Gray lines show FVS simulations. Blue lines show permanent plot data. Black lines show published yield curves. Red line shows median, and bands show 25th-75th and 10th-90th percentile range of FIA data.

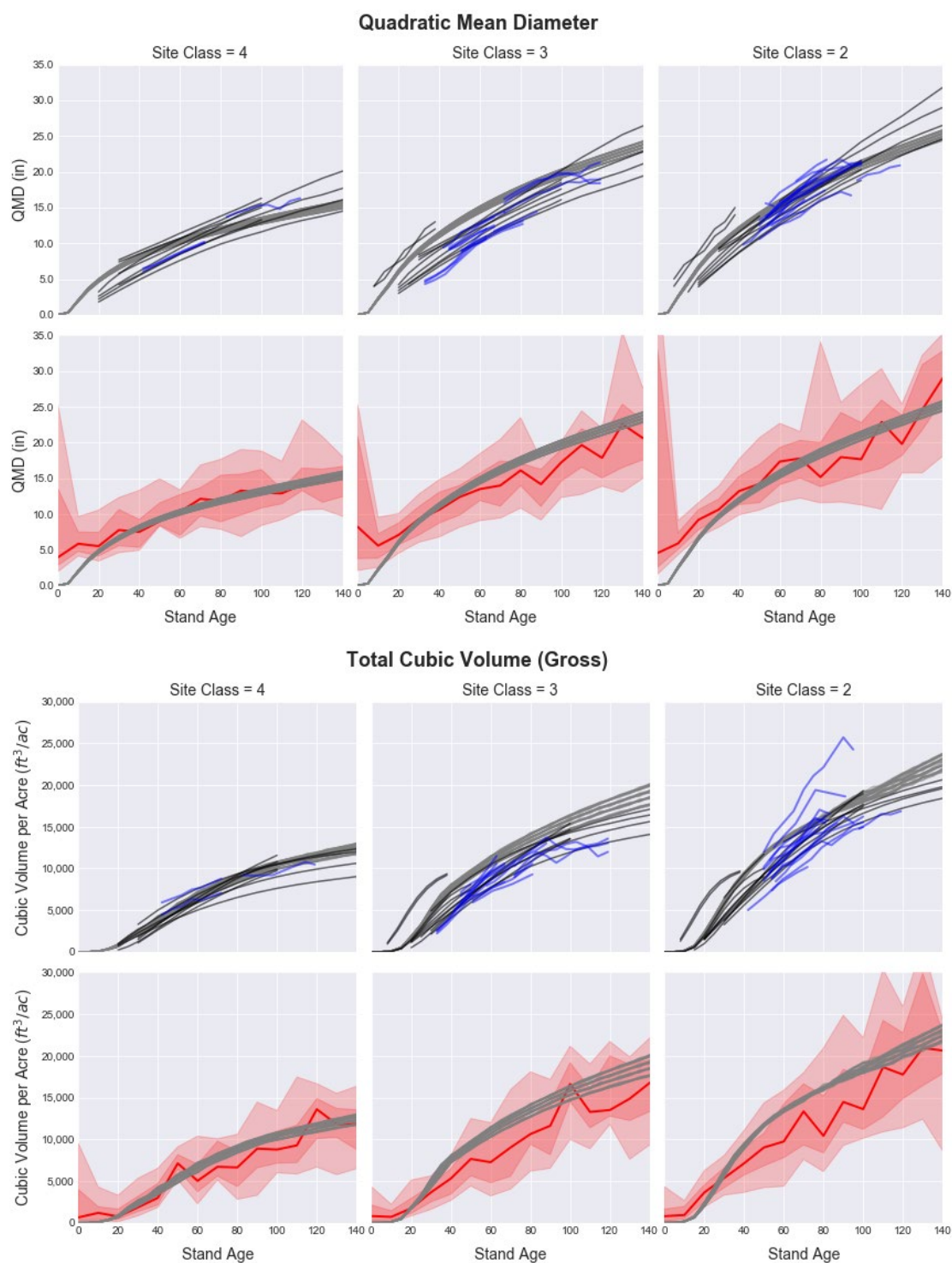


Figure S1.12. Modified FVS Parameterization (Part 2 of 3). Quadratic Mean Diameter (QMD, top 6 graphs) and Total Cubic Volume (bottom 6 graphs). Gray lines show FVS simulations. Blue lines show permanent plot data. Black lines show published yield curves. Red line shows median, and bands show 25th-75th and 10th-90th percentile range of FIA data.

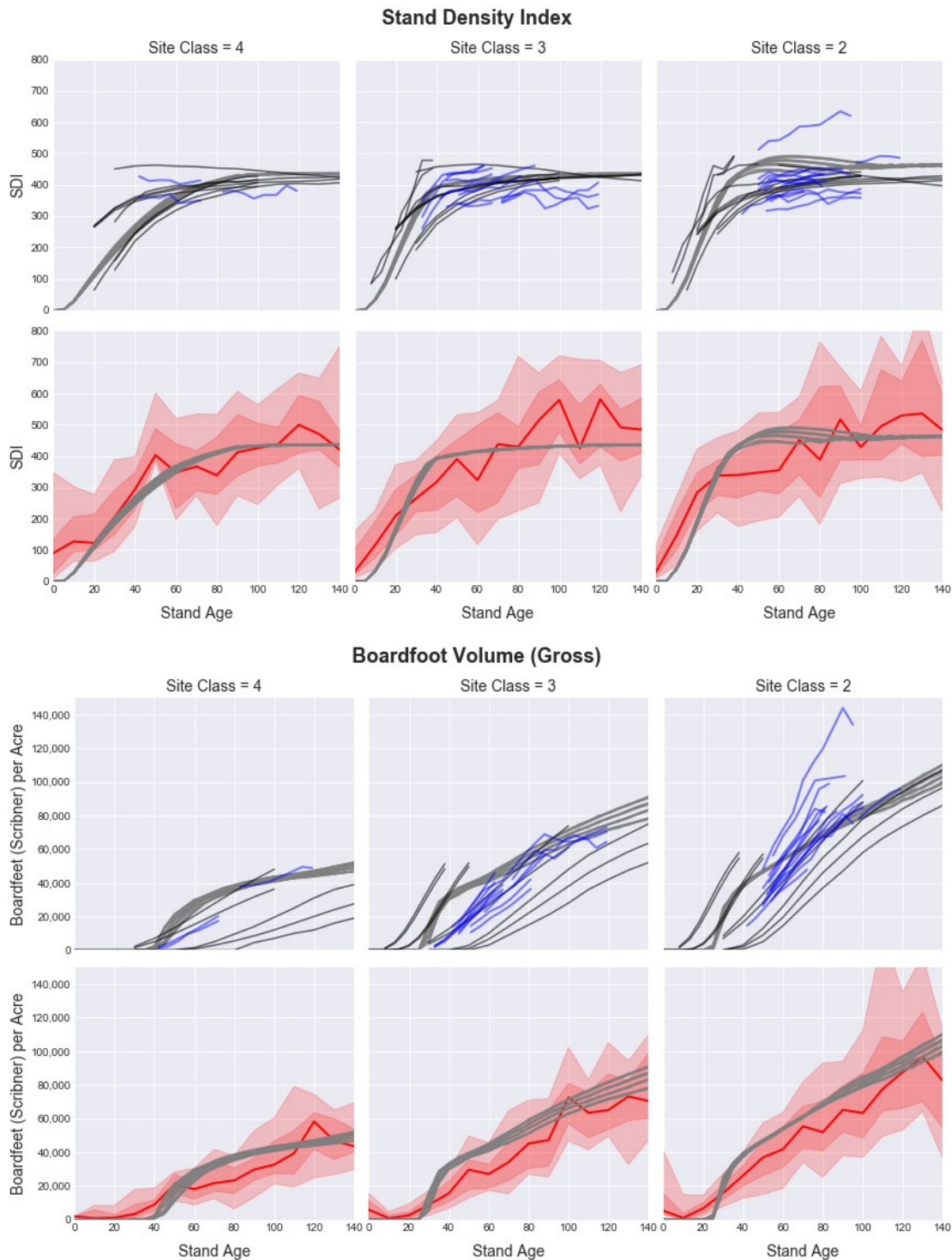


Figure S1.13. Modified FVS Parameterization (Part 3 of 3).

Stand Density Index (SDI, top 6 graphs) and Gross Boardfoot Volume (bottom 6 graphs). Gray lines show FVS simulations. Blue lines show permanent plot data. Black lines show published yield curves. Red line shows median, and bands show 25th-75th and 10th-90th percentile range of FIA data.

Table S1.8. Yield Table: Dominant Height (feet), 40 tallest trees per acre.

Age	Douglas-fir 50-year Site Index (King's)											
	75	80	85	90	95	100	105	110	115	120	125	130
20	37	38	40	42	44	46	48	49	51	53	55	57
25	44	46	49	51	54	56	59	61	63	65	68	70
30	51	54	57	60	63	66	68	71	74	77	80	82
35	58	61	64	67	71	74	78	81	84	88	91	93
40	63	67	71	74	78	82	86	90	93	97	100	103
45	68	73	77	80	85	89	94	98	102	106	109	112
50	73	78	82	85	91	96	100	105	109	114	117	120
55	78	82	86	89	97	102	107	112	116	121	124	127
60	82	87	91	93	102	108	113	118	123	127	131	134
65	85	91	95	97	107	113	118	124	128	133	137	140
70	89	94	98	100	112	117	123	129	134	138	142	145
75	92	97	101	103	116	122	128	134	139	144	147	150
80	95	100	104	106	120	126	132	138	144	148	152	155
85	97	103	107	109	123	129	136	142	148	153	157	160
90	100	106	110	112	126	133	140	146	152	157	161	164
95	102	108	112	114	129	136	143	150	155	161	165	168
100	104	110	115	116	132	139	146	153	159	164	169	172

Table S1.9. Yield Table: Trees per Acre

Age	Douglas-fir 50-year Site Index (King's)											
	75	80	85	90	95	100	105	110	115	120	125	130
20	368	368	368	368	368	368	368	368	368	368	368	368
25	368	368	368	368	368	368	368	368	367	367	367	367
30	368	368	368	368	368	368	367	367	367	367	367	366
35	365	365	364	363	357	355	353	350	357	355	353	337
40	362	361	360	359	345	334	323	314	335	330	323	306
45	359	358	356	355	306	296	288	276	313	305	295	277
50	356	355	353	350	277	268	258	251	292	281	269	251
55	354	351	349	345	254	245	238	229	271	259	246	228
60	351	348	343	336	235	226	220	213	252	238	225	208
65	347	340	335	327	218	211	204	198	234	220	207	191
70	341	333	327	317	204	198	192	186	217	204	191	176
75	334	326	318	308	192	186	180	175	203	190	177	170
80	328	318	310	299	182	176	171	166	190	177	165	158
85	321	311	302	290	173	168	162	158	178	166	155	151
90	314	304	293	281	165	160	155	150	168	156	149	146
95	306	293	283	273	158	153	148	144	159	147	144	140
100	294	283	273	264	152	147	142	138	150	142	137	134

Table S1.10. Yield Table: Basal Area (sq. ft. per acre)

Age	Douglas-fir 50-year Site Index (King's)											
	75	80	85	90	95	100	105	110	115	120	125	130
20	42	45	47	50	68	71	75	79	82	86	90	93
25	62	66	70	73	110	115	120	125	130	135	140	145
30	83	88	92	96	152	159	165	171	177	184	190	196
35	103	108	113	118	189	196	202	208	219	225	231	228
40	122	128	133	138	222	223	225	227	248	252	256	250
45	140	146	152	158	233	235	237	237	272	274	274	267
50	157	164	170	176	244	246	246	248	290	290	288	279
55	174	181	187	193	253	255	257	257	304	301	297	288
60	189	196	202	206	261	262	265	267	314	309	304	294
65	203	209	214	217	268	270	271	274	321	315	308	297
70	214	220	225	227	275	277	279	281	326	319	312	300
75	225	230	234	236	280	283	284	286	330	322	314	312
80	235	239	243	244	286	288	290	293	334	325	316	314
85	244	247	250	251	291	293	295	297	336	327	319	321
90	252	255	257	257	296	299	300	302	337	328	326	329
95	257	259	261	262	300	303	305	306	338	329	332	334
100	260	262	264	266	305	307	309	311	339	333	334	339

Table S1.11. Yield Table: Quadratic Mean Diameter (inches)

Age	Douglas-fir 50-year Site Index (King's)											
	75	80	85	90	95	100	105	110	115	120	125	130
20	4.6	4.7	4.9	5.0	5.8	6.0	6.1	6.3	6.4	6.5	6.7	6.8
25	5.6	5.7	5.9	6.0	7.4	7.6	7.7	7.9	8.0	8.2	8.4	8.5
30	6.4	6.6	6.8	6.9	8.7	8.9	9.1	9.3	9.4	9.6	9.8	9.9
35	7.2	7.4	7.5	7.7	9.9	10.1	10.2	10.4	10.6	10.8	11.0	11.1
40	7.9	8.0	8.2	8.4	10.9	11.1	11.3	11.5	11.6	11.9	12.0	12.3
45	8.5	8.7	8.9	9.0	11.8	12.0	12.3	12.5	12.6	12.8	13.1	13.3
50	9.0	9.2	9.4	9.6	12.7	13.0	13.2	13.5	13.5	13.7	14.0	14.3
55	9.5	9.7	9.9	10.1	13.5	13.8	14.1	14.3	14.3	14.6	14.9	15.2
60	9.9	10.2	10.4	10.6	14.3	14.6	14.9	15.2	15.1	15.4	15.7	16.1
65	10.4	10.6	10.8	11.0	15.0	15.3	15.6	15.9	15.9	16.2	16.5	16.9
70	10.7	11.0	11.2	11.5	15.7	16.0	16.3	16.6	16.6	16.9	17.3	17.7
75	11.1	11.4	11.6	11.9	16.4	16.7	17.0	17.3	17.3	17.6	18.0	18.4
80	11.5	11.7	12.0	12.2	17.0	17.3	17.6	18.0	17.9	18.3	18.7	19.1
85	11.8	12.1	12.3	12.6	17.6	17.9	18.3	18.6	18.6	19.0	19.4	19.7
90	12.1	12.4	12.7	12.9	18.1	18.5	18.8	19.2	19.2	19.6	20.0	20.3
95	12.4	12.7	13.0	13.3	18.7	19.0	19.4	19.8	19.8	20.2	20.6	20.9
100	12.7	13.0	13.3	13.6	19.2	19.6	19.9	20.3	20.4	20.8	21.2	21.5

Table S1.12. Yield Table: Cubic volume including top and stump (hundreds of cubic feet per acre)

Age	Douglas-fir 50-year Site Index (King's)											
	75	80	85	90	95	100	105	110	115	120	125	130
20	6.3	7.0	7.9	8.9	16.8	17.7	18.3	18.9	19.6	20.3	21.5	23.7
25	14.3	16.1	17.5	18.6	28.3	29.8	31.2	33.1	36.2	38.8	40.9	42.8
30	20.3	21.3	23.5	26.5	42.5	45.2	48.4	52.3	55.7	58.2	60.6	62.9
35	27.3	29.7	31.3	32.5	57.1	60.7	64.8	68.9	74.3	77.3	79.7	79.7
40	33.9	37.3	40.8	42.3	72.2	74.4	77.2	80.2	89.0	92.0	94.1	93.7
45	41.3	43.9	45.6	48.2	80.4	83.7	86.7	88.6	103.1	105.4	106.7	105.8
50	48.1	52.2	55.8	57.8	88.6	91.8	94.5	97.5	115.4	116.8	118.0	116.5
55	55.5	58.7	61.0	63.7	96.2	99.9	103.5	105.5	125.9	127.1	127.5	125.0
60	60.8	65.3	68.8	71.6	102.9	106.6	110.6	114.3	135.2	135.7	135.3	133.0
65	68.6	72.2	74.8	76.2	109.9	114.2	117.6	121.2	142.8	143.1	142.5	139.8
70	73.9	76.8	80.4	82.7	115.6	120.5	124.7	128.2	149.9	149.6	148.7	146.0
75	78.8	82.6	85.5	87.4	121.8	126.5	130.7	134.5	156.5	155.6	154.3	155.6
80	84.2	87.8	90.4	92.0	127.2	132.5	136.9	141.3	162.0	161.3	159.8	161.3
85	88.7	92.3	95.3	96.9	133.0	137.9	142.3	147.0	167.6	166.7	165.7	169.6
90	93.4	96.7	99.3	100.7	137.9	143.5	148.4	152.4	172.3	170.8	173.1	177.3
95	97.2	99.6	103.2	104.6	142.4	148.8	153.3	158.3	176.3	175.3	180.3	184.3
100	99.6	102.8	105.7	107.8	147.2	153.2	158.6	163.4	180.3	181.3	185.1	190.4

Table S1.13. Yield Table: Boardfoot Volume, Scribner Rule (thousand boardfeet per acre) *

Age	Douglas-fir 50-year Site Index (King's)											
	75	80	85	90	95	100	105	110	115	120	125	130
20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.9	1.3	1.9	2.9
30	0.0	0.0	0.0	0.0	5.4	9.5	12.8	16.0	18.1	20.7	23.2	25.7
35	0.0	0.0	0.0	0.0	22.6	24.9	27.1	28.7	30.8	32.2	33.5	33.2
40	0.9	1.8	2.0	2.4	29.9	30.4	32.1	33.1	37.3	38.5	39.4	39.1
45	3.7	5.1	8.8	13.0	33.5	34.6	36.3	37.1	42.7	43.3	43.9	43.0
50	11.5	15.2	18.8	21.2	36.9	37.8	38.5	39.8	46.9	47.0	47.2	46.3
55	18.4	21.0	23.8	25.6	39.1	40.2	41.3	42.0	50.3	50.4	50.6	49.8
60	23.2	25.7	28.2	29.8	40.8	42.5	43.7	45.0	54.0	54.1	54.3	53.6
65	28.0	29.7	31.3	32.0	43.5	45.0	46.5	48.4	57.5	57.6	57.9	57.1
70	30.6	32.3	33.6	34.7	46.0	48.0	50.2	51.8	61.1	61.4	61.2	60.8
75	33.2	34.7	35.9	37.0	48.6	51.0	53.1	55.5	64.8	64.8	64.8	66.1
80	35.5	36.6	38.1	38.6	51.3	54.3	56.6	59.2	67.8	67.9	68.0	69.4
85	37.3	38.5	39.6	40.1	54.3	57.2	59.9	62.2	71.0	71.1	71.7	73.6
90	38.9	40.0	41.1	41.3	57.3	60.7	63.2	65.3	73.8	74.1	75.5	78.2
95	40.0	41.2	41.9	42.7	60.2	63.6	65.6	68.5	76.6	76.6	79.2	81.4
100	40.8	42.1	42.8	43.6	63.0	66.1	68.6	71.3	79.0	80.1	82.3	84.0

* Scribner volume calculated for all trees $\geq 9"$ DBH to a minimum 6" top diameter inside bark, with a 1' stump.

Table S1.14. Yield Table: Stand Density Index

Age	Douglas-fir 50-year Site Index (King's)											
	75	80	85	90	95	100	105	110	115	120	125	130
20	105	110	116	121	154	160	167	173	179	186	192	198
25	144	151	158	164	227	235	243	252	259	267	276	283
30	181	189	197	204	295	305	314	324	333	343	352	361
35	215	224	232	240	349	358	367	374	392	400	408	400
40	246	255	264	272	394	393	393	394	428	433	436	424
45	274	284	293	301	400	400	401	397	455	455	452	438
50	301	311	320	328	406	406	404	404	472	468	462	444
55	325	335	344	352	411	411	411	409	483	475	466	447
60	347	357	365	369	416	414	415	416	489	478	466	446
65	367	374	380	383	418	418	417	418	490	478	463	443
70	382	388	394	395	422	421	421	421	490	475	460	440
75	396	401	405	405	423	424	422	423	488	472	456	450
80	408	412	415	413	425	426	425	426	485	469	452	446
85	418	421	423	420	428	427	426	427	482	465	450	450
90	427	429	429	426	429	430	428	428	478	461	455	456
95	433	431	432	430	430	431	430	429	474	456	458	457
100	433	433	433	433	432	432	431	432	470	458	455	459

Table S1.15. Yield Table: Curtis Relative Density

Age	Douglas-fir 50-year Site Index (King's)											
	75	80	85	90	95	100	105	110	115	120	125	130
20	19.8	20.7	21.5	22.4	28.2	29.2	30.4	31.4	32.4	33.5	34.7	35.7
25	26.5	27.6	28.7	29.7	40.5	41.7	43.1	44.5	45.7	47.1	48.4	49.6
30	32.7	34.1	35.4	36.6	51.6	53.2	54.8	56.3	57.8	59.4	60.9	62.3
35	38.4	39.8	41.1	42.4	60.2	61.7	63.1	64.3	67.2	68.5	69.8	68.3
40	43.5	45.0	46.4	47.8	67.3	67.1	66.8	66.9	72.7	73.3	73.7	71.5
45	48.2	49.7	51.2	52.5	67.8	67.6	67.6	66.8	76.6	76.4	75.9	73.2
50	52.5	54.1	55.5	56.8	68.3	68.3	67.7	67.6	78.9	78.1	76.9	73.8
55	56.4	58.0	59.4	60.6	68.8	68.6	68.4	67.9	80.2	78.7	77.1	73.8
60	59.9	61.5	62.6	63.2	69.1	68.7	68.7	68.6	80.7	78.7	76.6	73.2
65	63.1	64.1	65.1	65.3	69.2	69.1	68.6	68.6	80.6	78.3	75.8	72.4
70	65.4	66.3	67.1	67.2	69.4	69.1	69.0	68.9	80.1	77.6	74.9	71.4
75	67.5	68.2	68.7	68.6	69.3	69.3	68.9	68.8	79.4	76.7	73.9	72.8
80	69.3	69.8	70.2	69.8	69.4	69.3	69.0	69.0	78.8	75.9	72.9	71.9
85	70.9	71.2	71.3	70.7	69.5	69.3	69.1	68.9	77.9	75.0	72.4	72.3
90	72.3	72.3	72.2	71.4	69.6	69.5	69.1	68.9	77.0	74.0	72.9	73.0
95	73.0	72.5	72.5	72.0	69.5	69.5	69.2	68.9	76.1	73.1	73.3	73.0
100	72.9	72.5	72.5	72.2	69.6	69.5	69.2	69.1	75.2	73.1	72.6	73.1

Mapping Riparian Management Zone Buffers in Oregon and Washington

Constraints on harvest activities near streams are handled differently for each State's forest practice regulations (Oregon Secretary of State, n.d.; Washington State Legislature, n.d.), as well as under the FSC-US Standard for the Pacific Coast (FSC-US, 2012). The width of RMZs are displayed in Table S1.16.

Table S1.16. Buffer widths for Riparian Management Zones under FPA and FSC Rules (in feet)

Stream Type	Buffer Type	Oregon			Washington ¹			FSC-US Pacific Coast ³
		Stream Size			Site Class ²			
		SM	MED	LRG	II	III	IV	
Type F (fish-bearing)	core	20	20	20	50	50	50	50
	inner	30	50	80	78	55	33	150
	outer	--	--	--	42	35	27	--
Type SSBT (salmon, steelhead, bull trout)	core	20	20	--	--	--	--	50
	inner	20	30	--	--	--	--	150
	outer	20	30	--	--	--	--	--
Type N/Np (non-fish, perennial)	core	0	20	20	50 ⁴	50	50	25
	inner	0	30	50	0	0	0	75
	outer	--	--	--	0	0	0	--
Type Ns (non-fish, seasonal)	core	--	--	--	--	--	--	--
	inner	--	--	--	--	--	--	75
	outer	--	--	--	--	--	--	--

¹ Washington RMZ widths also differ whether stream bank-full width is above or below 10 feet. All buffers shown here, and all RMZs modeled in this study assume a > 10-foot bank-full width.

² Site Class categorizes forest growth potential using Douglas-fir 50-year Site Index (height of dominant trees at 50 years of age): II = 35.1-41.1 m (115-135 ft); III = 29.0-35.1 m (95-115 ft); IV = 22.9-29.0 m (75-95 ft)

³ FSC rules for the Pacific Coast do not identify a 'core' no-touch buffer zone, but rather distinguish inner and outer buffer zones for Category A (fish-bearing) and Category B (perennial non-fish-bearing) streams. FSC limits harvests to single-tree selection in inner buffers, and to single-tree or group selection in outer buffers (FSC-US, 2012). Because neither of these harvest systems are implemented in this study, all FSC buffer areas are modeled as no-touch (comparable to the core buffers of OR and WA ~FPA scenarios).

⁴ In Washington, Type Np (non-fish perennial) stream buffers are defined with additional criteria described below.

Geospatial layers identifying stream locations, stream types and site class were obtained from the Washington Department of Natural Resources, and a similar data layer for including stream locations, stream type and size was also obtained from the Oregon Department of Forestry. Buffers on streams were created using a Geographic Information System (GIS).

In Washington State, non-fish perennial stream (Type Np) RMZs are designated based on the location of sensitive sites and the stream's intersection with other streams. A 56-foot-radius buffer was assigned surrounding sensitive sites defined as a headwater spring or an intersection of two or more Type Np streams. Additionally, the length of 50 foot no-touch buffers applied to Np streams in cases where they intersect with a Type S (shorelines) or Type F (fish-bearing) stream are described in Table S1.17. The total area encumbered by no-touch and lighter-touch RMZ areas for FPA and FSC rules are visualized in Figure S1.14.

Table S1.17. RMZ length for 50-foot core (no-touch) buffer area for Type Np streams that intersect with Type S or Type F streams, Washington State.

Length of Type Np stream	Length of 50' Riparian Management Area
> 1,000'	500'
300' - 1,000'	50% of entire length or 300', whichever is greater
≥ 300'	The entire length

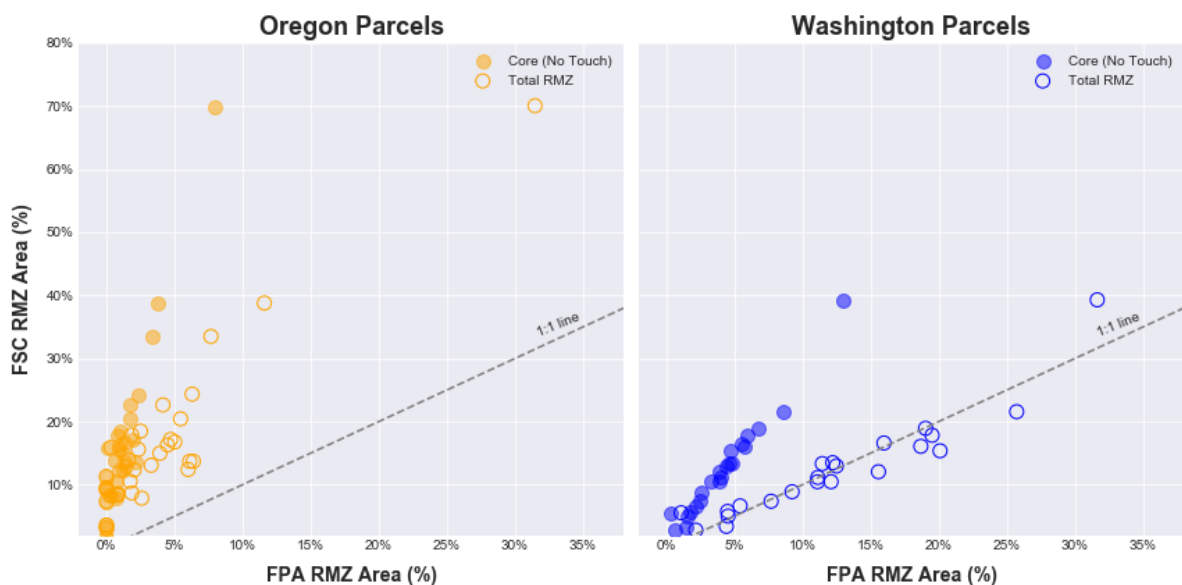


Figure S1.14. Comparisons of the area encumbered by FSC and FPA RMZs for Oregon and Washington.

Table S1.18. Minimum Retention for Inner RMZ Buffers on Type F streams, Washington State

Site Class	Harvest Age (years)	Green-Tree Retention * (trees per acre)
II	35	57
III	40	60
IV	40	66

** Retention is for conifer trees with DBH ≥ 12 , and must be at least 57 conifer TPA.*

All core RMZ buffer areas in Washington State are modeled as no-touch under the FPA scenarios. Only Type F streams have additional RMZ constraints beyond the core no-touch buffer. The outer RMZ buffer area in Washington requires the retention of at least 20 conifer TPA with DBH ≥ 20 ". For the inner RMZ buffer area, we used an online tool produced by the Washington Department of Natural Resources (Washington State Department of Natural Resources, 2009) to estimate the green-tree retention during a proposed harvest to meet a desired future condition for the stocking of conifer trees of 325 sq. ft. per acre at age 140. With this tool, we identified the minimum retention requirements for a "thin from below" treatment in the Inner RMZ buffer area for Type F streams.

Silvicultural Prescriptions for RMZs in Oregon under FPA Rules

All core RMZ buffer areas in Oregon are modeled as no-touch under the FPA scenarios. Retention requirements in Oregon FPA rules are defined in terms of retention for a pre-defined length of stream (e.g., 1,000 ft). These retention requirements were converted to trees per acre and basal area per acre by dividing the specified retention by the area that would fall within each buffer type for the pre-defined length of stream.

In practice, these retention requirements are enforced at the level of the entire RMZ rather than for separate RMZ buffer areas. To simplify growth-and-yield simulations, we model each RMZ buffer area independently and do not count retention in other RMZ buffers towards a total value. This approach does not follow the common practice of "packing" retention into core and inner buffer areas, and thus is likely to represent a higher level of green tree retention than is required by Oregon FPA rules. For example, we model all core RMZ buffers as no-touch, and

require all inner RMZ buffers to meet the per-acre retention requirements without consideration of the retention neighboring core buffers.

For Type F and Type N streams, Oregon FPA rules following a branching logic approach. Each stream size in these types has a pre-defined “Standard Target.” If, at the time of the proposed harvest, a stand is at or above the Basal Area Standard Target, the retention must meet the levels shown in Table S1.19. If a stand falls between 50-100% of the Basal Area Standard Target, the retention must meet the levels shown in Table S1.20. If the stand does not currently meet or exceed these retention levels, no harvest is permitted. Harvests within RMZ buffers are modeled as a “thin from below,” removing the smallest trees from the stand’s diameter distribution first, while enforcing the minimum TPA and DBH requirements for retention.

Table S1.19. Oregon FPA “Standard Targets” for Retention in Inner RMZ Buffers

Retention Type	Type F			Type N			Type SSBT ¹	
	SM	MED	LG	SM	MED	LG	SM	MED
Basal Area ² (sq. ft. per acre)	35	75	100	-- ⁴	44	56	81	100
Trees per Acre ³	-- ⁵	19	18	--	9	19	18	22
Minimum DBH ³ (inches)	--	8	11	--	8	11	8	8

¹ Type SSBT Streams have only one set of retention targets, they do not another retention option as is the case with Type F and Type N streams. If a stand does not meet SSBT targets at the time of proposed harvest, no harvest is permitted.

² FPA rules for Type F and Type N streams count basal area in live conifer trees with DBH ≥ 6 ". For Type SSBT streams, the basal area target may include any species of trees with DBH ≥ 6 "

³ The requirements for the minimum number (TPA) and size (DBH) of retained trees refers to live conifers only. Although exceptions exist in Oregon FPA rules for different stream types and sizes, snags and hardwood trees are not counted toward retention requirements as modeled in this study.

⁴ Small Type N streams do not have specific retention requirements for inner RMZ buffers. These zones are modeled the same way as non-riparian areas.

⁵ Small Type F streams have no specific tree size or count requirements for inner RMZ buffers.

Table S1.20. Oregon FPA “Alternative Targets” for Retention in Inner RMZ Buffers

Retention Type	Type F			Type N		
	<i>SM</i>	<i>MED</i>	<i>LG</i>	<i>SM</i>	<i>MED</i>	<i>LG</i>
Trees per Acre ¹	61	63	66	-- ²	61	63
Minimum DBH ¹ (inches)	6	6	6	--	6	6

¹ The requirements for the minimum number (TPA) and size (DBH) of retained trees refers to live conifers only.

² Small Type N streams do not have any specific retention requirements for inner RMZ buffers. These zones are modeled the same way as non-riparian areas.

Chapter 2. Predicting forest structure attributes from lidar or satellite data and assessing model transferability across multiple ecoregions

2.1 INTRODUCTION

For natural resource scientists, agencies, managers, and service providers who work across forested areas at the scale of ecoregions or states/provinces, the production of wall-to-wall maps of forest conditions that can be updated on at least an annual basis remains a common pursuit. The production of such maps has been aided by the rapidly expanding availability of remotely sensed data used to characterize forest conditions across landscapes and to monitor and attribute changes in them over time. In the field of forest science, data from airborne laser scanning (also known as lidar) has taken on a particularly prominent role for addressing both commercial forest management and ecological concerns. In fact, several US states and European countries have begun investing in periodic province-scale airborne lidar acquisitions; however, large areas of the world have no airborne lidar available, or partial or incomplete coverage. The emergence of satellite constellations that release high-resolution imagery to the public with a high frequency provides a clear opportunity to update predictive models and monitor various conditions on the earth's surface across entire regions frequently. For example, the Sentinel-2 mission employs two satellites to offer imaging in the visible and near-infrared spectra at 10-meter resolution every 5 days since 2015. The increasing accessibility of satellite imagery through cloud services such as Google Earth Engine (Gorelick et al., 2017) is also substantially enhancing the capacity and usage of these data.

Systems leveraging remote sensing data with pre-existing timber cruise and stand examination data for stand- and watershed-level planning have also emerged, such as the FSVeg Spatial Data Analyzer used within the USFS to make maps with imputed plot data based on correlational models trained using Landsat imagery and Common Stand Exam field data (Arciniega, 2017; Haberland and Marston, 2017; USDA Forest Service, Natural Resource Manager, 2021). The primary bottleneck, as in many computer vision and Machine Learning (ML)

challenges, remains the limited volume of confidently located and labeled training data samples which can be co-registered with remote sensing data layers.

In the Pacific Northwest, all state and federal agencies with significant involvement in active forest management now employ remote sensing forest inventory systems, including substantial investments into targeted field campaigns to collect forest data based on coincidence with lidar acquisitions (Hudak et al., 2014) and digital aerial photogrammetry (Strunk et al., 2019). Although all these natural resource agencies share generally similar interests in mapping out forest attributes including canopy cover, tree height, stem density, volume, and biomass/carbon, they continue to design and implement field campaigns with widely varying sampling strategies and plot designs. The compilation of these data into a larger training dataset has provided an important step towards the creation of a diverse regional dataset of confidently located labeled training data for diverse forests across the Pacific Northwest.

Investigations into the transferability of remote sensing predictive models across space and time

Generating maps covering large landscapes and involving predictive models trained on multiple data sources naturally leads to questions regarding the degree to which models trained in one region or time can be transferred to others. Numerous studies have investigated the potential transferability of models that predict forest attributes from remote sensing to other regions and times (Foody et al., 2003; Cutler et al., 2012; Fekety et al., 2015, 2018; Hooper and Kennedy, 2018; Jin et al., 2018; Luther et al., 2019; Tompalski et al., 2019; Ewijk et al., 2020; Marino et al., 2022). These investigations have generally been performed without the creation of public access to the underlying data, and appear to be much more commonly applied to lidar data as opposed to satellite imagery.

In contrast to most of the earlier studies into transferability of remote sensing models for forestry, this study relied exclusively on data collected by other organizations and intended for their own internal uses. I expect the comparatively large sample size I was able to secure by compiling datasets that are technically in the public domain but are not publicly available should be an indicator of a wave of additional inventory data that can be released over time and that this diverse mixture of sampling protocols and data sources is likely to become a more common situation for forest mapping efforts into the future. Among the previous transferability studies that used forest measurements collected on the ground, most have relied upon relatively small research

datasets (e.g., 100-500 plots) from a relatively small number (e.g., 2-3) of ecoregions (see Table 2.).

Previous research into the transferability of lidar-based forest structure models has indicated that transferring models between regions and across lidar acquisitions may lead to modest decreases in predictive accuracy and increased model bias, but that relationships between predictor and target variables are largely conserved and model transfer remains a viable use case (Fekety et al., 2015, 2018; Tompalski et al., 2019). In addition, Fekety et al. (2015) reported that the inclusion of additional co-registered samples of single field plots with multiple instances of remote sensing from different times can improve these models.

The performance of predictive models is largely sensitive to the algorithm chosen, and may correspond to different generalization capabilities for model transfer among regions. Marino et al. (2022) documented the transferability of parametric models (linear and exponential) between lidar acquisitions to be dramatically affected by lidar dataset qualities such as pulse density, and can produce significantly biased estimates of forest structure when transferred. Cosenza et al. (2020) compared linear models fit using Ordinary Least Squares (OLS) regression, Random Forests (RF), and k-Nearest Neighbor (kNN) models in five different countries spanning tropical to boreal forests and found that OLS and RF models generally showed comparable performance, and that kNN consistently performed the poorest. Fekety et al. (2015, 2018) utilized RF models across multiple lidar acquisitions and study areas in northern Idaho and concluded these models generalized relationships between forest structure and lidar attributes well.

The transferability of remote sensing models can also vary among target variables predicted. Fekety et al. (2018) demonstrated that lidar based RF models show better transferability for predictions of basal area (BA) than they do for stem density, with mean-Normalized Root Mean Squared Error (Normalized Root Mean Square Error, NRMSE) for withheld field plots ranging 32-51% for BA and 40-67% for stem density. Fekety and colleagues (2018) concluded that their research “supports extending lidar-based prediction models to lidar extents that do not have tree data, assuming the structural characteristics of the unmeasured forest were included in the original predictive model but measured elsewhere within the ecoregion.”

Table 2.1. Studies investigating the transferability of forest structure models trained with remote sensing data

Study	# Plots	# Ecoregions	Locations	Algorithms
Foody et al. (2003)	144	3	Brazil, Thailand, Malaysia	MLP
Cutler et al. (2012)	144	3	Brazil, Thailand, Malaysia	MLP
Fekety et al. (2015)	173	1	Idaho, USA	kNN, RF
Fekety et al. (2018)	360	1	Idaho, USA	RF
Luther et al. (2019)	97	1	Newfoundland, Canada	OLS, kNN, RF
Tompalksi et al. (2019)	239	1	Coastal British Columbia, Canada	OLS, kNN, RF
Ewijk et al. (2020)	479	2	Alberta and Ontario/Quebec, Canada	OLS, RF
Mauro et al. (2021)	1,133	6	Oregon, USA	OLS, RF
Marino et al. (2022)	256	1	Segovia, Spain	OLS
This study	2,550/4,199*	9	Oregon and Washington, USA	OLS, kNN, RF, GBM

Notes: Abbreviations used for algorithms include: MLP = Multi-Layer Perceptron; kNN = k-Nearest Neighbors; RF = Random Forest; OLS = Ordinary Least Squares (linear models including exponential and power transformed linear models); and GBM = Gradient Boosting Machines.

** Sample size for this study shows 2,550 unique plots co-registered with lidar, and 4,199 unique plots co-registered with satellite imagery. When including grown-forward plot observations intersecting with multiple years of lidar scanning or satellite imagery, sample size grows to 2,739 lidar plots and 7,816 satellite plots.*

When working across larger geographic areas, Hooper and Kennedy (2018) advocated for the adoption of an ensembling approach that combines multiple locally-trained models fit to spatially-overlapping subsets of the global dataset. Making predictions of land cover and impervious surface presence across Washington, Oregon, and California, Hooper and Kennedy reported a “subtle” effect where model performance decreased as the extent of the training dataset used for each member of the ensemble increased, although this effect was not consistently seen for predictions of forest cover and impervious surface cover, the two cases they investigated. When lidar-based models for predicting forest structure were transferred from western to eastern Canada boreal forests, Ewijk and colleagues (2020) found including data from the target region in training data for a model substantially improved predictions, and that the best transferred models performed within $\pm 5\%$ NRMSE of the locally-trained model, but showed increased relative bias by up to 10% (i.e., average bias normalized by the mean response for each target variable).

Mauro and colleagues (2021) demonstrated an approach to adapting Machine Learning (ML) models like Nearest Neighbors algorithms and Random Forests to semi-parametric forms that allow “calibration” in new areas by adding a random effects term along with the non-parametric model predictions to capture variation at plot- and lidar-acquisition scales not explained by the non-parametric model. However, they reported “The minor effect that calibration had in the semiparametric models can be explained by the ability of random forest to effectively use large numbers of predictors and model nonlinear patterns. If the training dataset covers the main ecological gradients of the area under analysis, random forest captures the most important sources of variability that can be explained by the auxiliary information leaving little room for improvements in the calibration step.”

Although the growing use of remote sensing data is increasingly influencing how and when field observations are collected, most modeling efforts that seek to incorporate forest inventory data from multiple locations and years, or using multiple sources of remote sensing (e.g., satellite and lidar) commonly run into issues that the timing of field observations may differ from the timing of remote sensing observations by months or years. This issue is particularly relevant when remote sensing data are available with a relatively high frequency compared to field measurements which are not commonly repeated in the same location over time, and in light of the fact that repeated remote sensing observations of the same inventory plot can improve model performance (Fekety et al., 2015). To effectively leverage the information that can be gleaned from costly field

measurements of forest conditions, it is common to estimate forest conditions at the time of remote sensing observations based on the inventory data and some model of forest change over time (de Lera Garrido et al., 2020).

While the development of modeled observations derived from forest inventory plot measurements can significantly increase the sample size available for model training and validation due to repeated remote sensing (see Table 2.2), this practice can also introduce systematic biases in the modeling effort. De Lera Garrido and colleagues (2020) highlighted the potential for systematic errors based on underlying growth model biases, errors in excluding plots that have been disturbed prior to remote sensing, and drift between how well the modeled data distribution relates to the location where it is applied (e.g., as modeled forest conditions diverge from the structure and composition in the field).

Spatial correlation within datasets has also been shown to be a source of information leakage that violates assumptions of independence between cross-validation folds and/or train/test splits (Hijmans, 2012; Schratz et al., 2019). In a study into the transferability of tropical forest biomass models, Cosenza et al. (2020) suggested that five-fold cross-validation was more likely to avoid issues of over-fitting that might lead to biased estimates of model performance when leave-one-out cross validation is employed. In classification and regression problems across diverse domains, it has repeatedly been observed that “flat” cross-validation that is also used to perform feature selection and hyperparameter tuning can provide optimistically-biased estimates of generalization ability that can be reduced with nested cross-validation (Varma and Simon, 2006; Krstajic et al., 2014; Tsamardinos et al., 2015). That being said, Wainer and Cawley (2018) have since suggested that a comparison of flat and nested cross-validations schemes in the case of binary classification problems shows nested cross-validation schemes may commonly be overzealous.

Unmet demand for data and technology transfer to support fundamental forest assessment and planning tasks

The effort described in this study to produce wall-to-wall forest maps in Oregon and Washington was originally conceived within the scope of a collaborative project to develop actionable science and decision-support tools for non-industrial forest owners (Diaz et al., 2020). One of the objectives in this project was to develop a reproducible workflow to produce maps that could be integrated into an open-source web application called Landmapper

(<https://landmapper.ecotrust.org>) to reduce the time and effort required for the completion of forest stewardship plans by non-industrial forest owners and their service providers.

In this chapter, I have summarized findings and lessons learned from an effort to identify and deploy models suitable for statewide mapping of forest structure across Oregon and Washington, USA with the objective of providing maps of forest attributes for use at stand- and property-scales (i.e., less than 100 acres). I have reviewed how well different algorithms and different types of remote sensing data (lidar vs. satellite) perform in predicting forest structure attributes.

I have characterized how well these models appear to generalize the relationships between remote sensing features and forest structure attributes and test which algorithms, once tuned, most accurately predict six forest structure attributes using indicators for model precision and bias across nine ecoregions across the states of Oregon and Washington, USA. I have quantified how much each model's performance is affected by increasing levels of access to information across a landscape and a target region and report how model performance and transferability is affected by the inclusion of sample weighting based on prior knowledge of forest conditions in the target region.

I implemented several approaches to try and mitigate the potential biases introduced by using modeled estimates of forest conditions to build predictive models, including: using modeled estimates that are no more than 10 years from the date of measurement; eliminating plots from the training dataset once they appear to have been disturbed; and employing nearest-neighbor based re-weighting of sampling data to try to control for the “drift” or domain adaptation issue of fitting models to data that are distributed differently from the population the models will be applied in. I explored a similar process to the calibration approach employed by Mauro and colleagues (2021), utilizing sample weighting based on prior information about the target region.

In response to previously-documented concerns regarding the practice of cross-validation, I ensured that all repeated observations of individual plots were kept within a single fold for either training or testing of models, although I did not adopt folding strategies that separate plots from geographically similar areas into distinct folds. In addition, the regional hold-out strategies I employed in fitting “Visitor” and “Outsider” models also helped to inform the effects on predictive performance when models are trained without data from the region in which they are applied.

Primary research questions/hypotheses

At a basic level, I expected the larger sample size compiled for this study compared to most previous studies that have investigated transferability of regional forest structure models will correspond to improved model performance and reduced bias compared to previous reports. I expected linear models to show the weakest capacity for transferring effectively between regions, and RF and GBM models to show the strongest ability to generalize and transfer between regions. Given the nature of many ML algorithms as “data-hungry”, I expected RF and GBM models to show improved performance in predicting forest conditions in target ecoregions based on the inclusion of training from outside those ecoregions.

With the volume and diversity of training data available in this study, I expected variation observed when transferring models trained in one region to another region to be dramatically reduced when these “Visitor” models are compared to models trained with data spanning multiple ecoregions. Across the varying scopes of training data described in the regional “hold-out” scopes described in more detail in Methods below, I expected model performance to improve following the general pattern of Visitor < Outsider < Insider < Global models. Finally, I expected sample weighting as a strategy for domain adaptation would substantially improve the performance of predictive models, particularly in regions where the distributions of forest structure attributes in training and testing sets are more disparate.

Entering this study, I expected in general that lidar-based models would consistently outperform satellite-based models, based on the higher density of information collected by lidar and closer correspondence of active (versus passive) remote sensing to the underlying forest structure that I sought to characterize and predict using six different attributes. While lidar models would be preferable, I expected satellite models would offer predictive performance that is sufficiently precise and unbiased to satisfy needs for property-scale planning (i.e., from dozens to hundreds of acres) in extensive management scenarios and for regional-scale modeling and monitoring that can be consistently applied for annually updated wall-to-wall maps. In use cases involving intensive forest management where precise estimation of timber volume and value is a primary concern, I expected satellite models will provide insufficient precision.

I expected ML models such as k-Nearest Neighbors (kNN), Random Forests (RF), and Gradient Boosting Machines (GBM) to outperform linear models in terms of model precision and to show less bias across the range of each attribute predicted. I expected sample weighting during

model training would reduce biases for predictions in forest conditions that are underrepresented in the training data relative to forest conditions across each ecoregion (e.g., lower canopy cover and lower biomass forests).

2.2 METHODS

Compilation of multi-agency remote sensing plot data

Data from several thousand field plots were compiled spanning nine ecoregions across Oregon and Washington states in the Pacific Northwest, USA (see Figure 2.1 and Table 2.2). Ground-truth data (forest plots) were gathered from public records requests and direct communications with agency staff. These records are from field campaigns implemented from 2010-2018 by the US Forest Service (USFS), Bureau of Land Management (BLM), and Washington Department of Natural Resources (DNR).

Each of these campaigns adopted distinct plot designs, but all were intended for the explicit purpose of developing models that correlate remote sensing data with forest attributes to build predictive models. The USFS utilized the largest sampling plots, covering 1/4-acre each (58.9 ft radius), while BLM utilized 1/8-acre plots (41.6 ft radius), and DNR utilized 1/10-acre plots (37.2 ft radius). Plots from each agency are limited to their own forest ownerships in the region, corresponding to a concentration of BLM plots in southwestern Oregon, for USFS plots along the Cascade Range and at higher elevations in the Blue Mountains, and with DNR plots scattered broadly across Washington state, but generally at lower elevations in the Cascade Range compared to USFS sample locations (see Figure 2.1).

These plot samples span significant climatic and biophysical gradients from the Pacific Coast to the inland Northwest. They include samples from even-aged and multi- or uneven-aged forests, including both naturally regenerated stands and tree plantations. Several of the major differences between forest structure across ecoregions can be observed in the distributions of six target variables recorded in the USFS Forest Inventory & Analysis (FIA) plot network illustrated in Figure 2.2. In particular, forests in ecoregions west of the Cascade Crest tend to be denser in terms of canopy cover and stem counts, with taller and larger diameter trees than are found in the drier ecoregions east of the Cascade Crest.

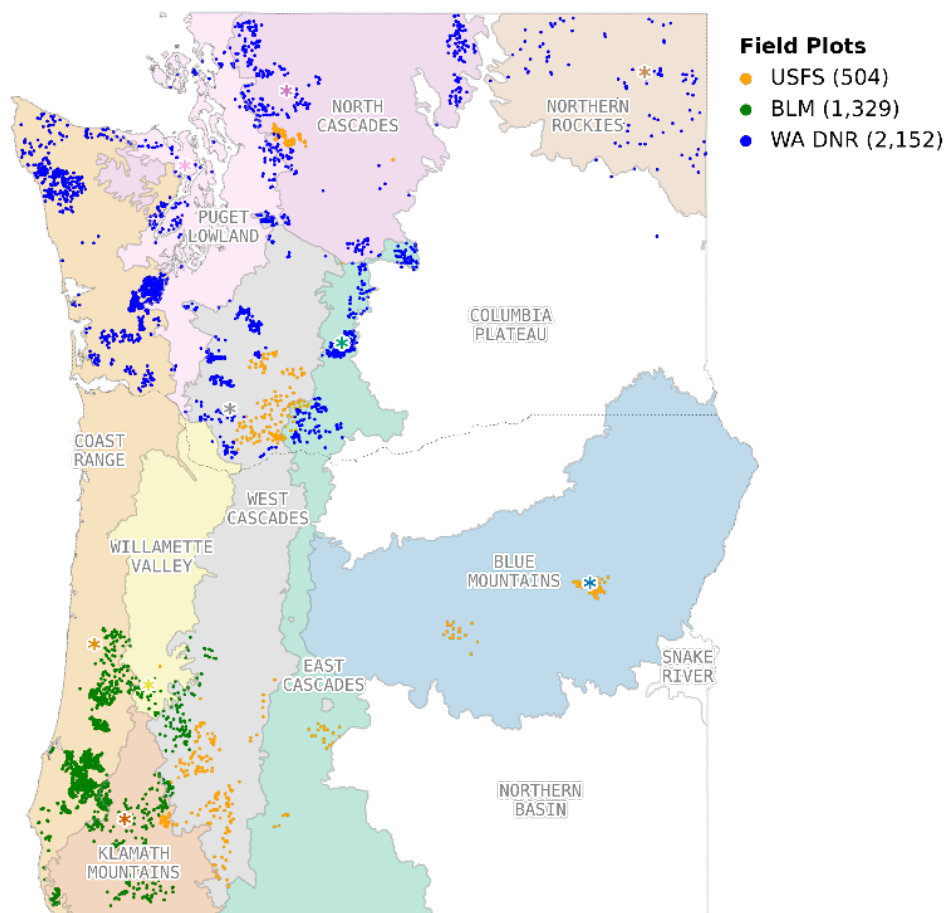


Figure 2.1. Locations of remote sensing forest plots.

Plots measured by Washington Department of Natural Resources (WA DNR), USDA Forest Service (USFS) and USDOJ Bureau of Land Management (BLM) across Oregon and Washington states, USA. Names of EPA Level III Ecoregions are labeled.

Table 2.2 Plot-level observations used for model training and evaluation from each year of lidar acquisitions and satellite imagery across Pacific Northwest ecoregions.

ECOREGION	LIDAR YEARS									SATELLITE YEARS			
	Unique Plots	2012	2013	2014	2015	2016	2017	2018	Total	Unique Plots	2019	2020	Total
Blue Mountains	108	--	--	17	91	--	--	--	108	222	219	217	436
Western Cascades	545	--	69	21	22	193	170	104	579	952	937	914	1,851
Coast Range	879	4	291	192	75	129	208	57	956	1,485	1,435	1,117	2,552
Eastern Cascades	191	--	--	44	30	61	3	53	191	556	548	544	1,092
Klamath Mountains	183	--	162	--	8	26	6	--	202	231	224	191	415
North Cascades	382	--	--	11	43	218	86	49	407	448	440	437	877
Northern Rockies	108	--	--	--	6	38	32	42	118	111	108	108	216
Puget Lowland	111	--	2	10	42	41	22	9	126	149	145	145	290
Willamette Valley	43	--	--	--	--	1	37	5	43	45	43	44	87
Total	2,550	4	524	299	317	709	564	322	2,739	4,199	4,099	3,717	7,816

Notes: Individual field plots may appear in more than one year of lidar or satellite data. During model fitting and cross-validation, all observations of a given plot were kept in the same cross-validation fold (i.e., used only for training or testing).

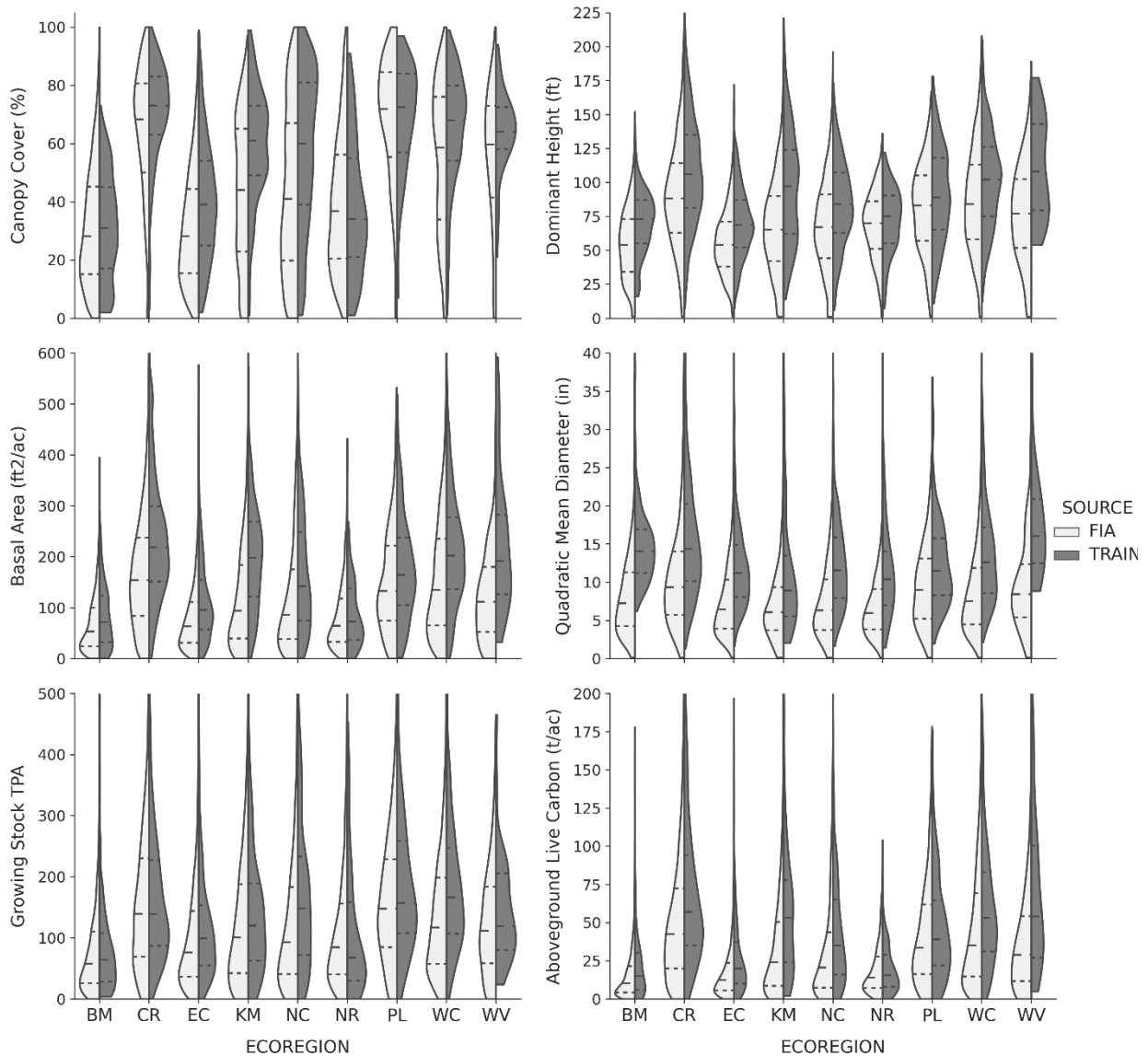


Figure 2.2. Distributions of forest structure variables across ecoregions comparing training datasets with all forested areas in Oregon and Washington.

The distributions of the six target variables are displayed for each ecoregion. Each violin in each graph displays the probability density of each variable in observations from the FIA dataset (unshaded, left) and remote sensing training dataset (dark gray, right). The lines inside each violin show the 25th, 50th, and 75th percentiles of each distribution. Ecoregion abbreviations are: BM = Blue Mountains; CR = Coast Range; EC = East Cascades; KM = Klamath Mountains; NC = North Cascades; NR = Northern Rockies; PL = Puget Lowlands; WC = West Cascades; WV = Willamette Valley. These distributions are drawn from all forested FIA “conditions” with each distinct condition given equal weight. Values for plot-level attributes were calculated using FVS from FVS-ready FIA data (Shaw and Gagnon, 2020a).

Compiling remote sensing data from publicly-available sources

Lidar and satellite imagery were gathered and processed into summary metrics for each inventory plot. These remote sensing data were compiled to offer a single set of plot-level predictor variables for each plot in each year where remote sensing data were available. Model-ready training datasets for satellite-based and lidar-based models have been published online¹.

Lidar data were initially extracted from 111 different lidar acquisitions conducted between the year 2000 and 2018. Lidar point clouds were primarily accessed from NOAA Digital Coast², the Puget Sound Lidar Consortium website³, the Washington DNR Lidar Portal⁴, and from an FTP server for lidar data hosted by Oregon State University⁵. A small number of lidar datasets not available from these public sources at that time were acquired through direct coordination with USFS staff (Bob McGaughey, personal communication). Each of the lidar acquisitions employed varying instrumentation and collection procedures, producing point clouds of varying densities, particularly when comparing earlier acquisitions to more recent acquisitions. Point densities among plot clips ranged from a low of 1.2 points per square meter to a high of 183.4 points per square meter; 95% of plots had a lidar point density higher than 4.98 points per square meter. The median point density among plots was 17.7 points per square meter.

At each inventory plot location, a 1-hectare circular clip of the point cloud was extracted using the LAStools lasclip command line tool. One-hectare point cloud clips encompassing each inventory were processed to generate ground surfaces using the LAStools lasground command line tool, and then clipped to the fixed radius plot where tree measurements were collected (ranging from 1/10-acre to 1/4-acre plots depending upon the agency that collected the field data). Point clouds from these smaller plot clips were then summarized to the plot-level using the FUSION lidar software CanopyModel, CloudMetrics and GridSurfaceStats command line tools. Grid-based calculations (e.g., percent canopy cover) using lidar data were calculated using a 0.5-meter cell size.

¹ Satellite-based training data are here:

https://github.com/ECOTRUST/ForestMapping/blob/master/data/processed/sentinel_structure_training_data.csv

Lidar-based training data are here:

https://github.com/ECOTRUST/ForestMapping/blob/master/data/processed/lidar_structure_training_data.csv

² <https://coast.noaa.gov/digitalcoast/>

³ <https://pugetsoundlidar.ess.washington.edu/lidardata/>

⁴ <https://lidarportal.dnr.wa.gov/>

⁵ <ftp://lidar.engr.oregonstate.edu>

Satellite data were extracted and summarized covering the footprint of each inventory plot for the years 2019 and 2020 using Sentinel-2 surface reflectance imagery (at 10-meter resolution for most bands) available from Google Earth Engine⁶. Sentinel-2 imagery covering plot footprints using 10 bands (blue, green, red, near-infrared, two shortwave infrared bands, and four red edge bands) were medoid-composited into leaf-on (April 1 through Sept 30) and leaf-off (October 1 of preceding year through March 31 of current year) periods. Plot-level averages of these 10 bands were calculated and then used to generate several derived indicators including Normalized Difference Vegetation Index (NDVI) (Tucker, 1979; NASA, 2000), Soil Adjusted Vegetation Index (SAVI) (Huete, 1988; US Geological Survey, 2022), and tasseled cap brightness, greenness, and wetness (Shi and Xu, 2019). I also calculated the differences for each of these spectral values and derived indices between leaf-on and leaf-off periods. The full suite of predictor variables used from lidar and satellite data sources are shown in Table 2.3.

To help identify plots that had been disturbed between the time of their field measurement and the time of lidar or satellite observation, I utilized the Landtrendr change detection algorithm (Kennedy et al., 2010) based on short-wave infrared (SWIR) and normalized burn ratio (NBR) indices calculated from Landsat imagery from 1984 to the year of lidar or Sentinel-2 observation. Locations where a disturbance was detected using either the SWIR or NBR index between the year of field measurement and the year of lidar or Sentinel-2 observation were excluded from subsequent modeling. Landtrendr modeling was implemented on Google Earth Engine (Kennedy et al., 2018).

Modeling plots forward in time to align with remote sensing data

To join remotely sensed data with field observations, field plots were projected to the year of remote sensing using regional variants of the Forest Vegetation Simulator (FVS). FVS was used to grow forward plot-level tree lists forward in time using an annual time-step. The differences between the original measurement years and remote sensing data collection years are presented in Table 2.4. In areas where multiple lidar acquisitions had been collected over time, each separate year of lidar data was incorporated as a distinct sample using the closest year's observed or simulated tree data.

Table 2.3. Predictor variables used in lidar-based and satellite-based predictive models.

⁶ https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR

Data Source	Predictor Variables
<i>Lidar</i>	17 lidar-specific variables, including ⁷ : • Proportion of lidar returns within 1 of 6 vertical strata (0.15 – 1.37, 1.37-5, 5-10, 10-20, 20-30, and 30+ meters above ground) • 5th, 25th, 50th, 75th, and 95th percentile of height above ground • standard deviation of height above ground • % canopy cover • potential volume and surface volume • kurtosis and skewness
<i>Satellite</i>	53 satellite-specific variables, including: 45 Sentinel-2 variables based on 10 spectral bands (R, G, B, NIR, SWIR2, SWIR2, RE1, RE2, RE3, and RE4): • Surface reflectance in 10 spectral bands during leaf-on and leaf-off periods • NDVI, SAVI, and tasseled cap brightness, greenness, and wetness • Change in surface reflectance for each of 10 bands and 5 derived indices between leaf-on and leaf-off period 8 Landtreindr variables derived from Landsat 7 and Landsat 8 imagery: • Duration, magnitude, rate, and years since last disturbance calculated based on SWIR1 Landsat band. • Duration, magnitude, rate, and years since last disturbance calculated based on Normalized Burn Ratio calculated from Landsat SWIR1 and NIR bands.
<i>All</i>	Elevation, Latitude, Longitude

Although several of the survey campaigns from which I sourced these plot data have adopted manual inspection procedures for quality assessment and quality control, a substantial number of plots still appear to be poorly co-registered or to have been disturbed sometime between the date of field measurement and the observation date via remote sensing. In addition to excluding measured or modeled plot data where disturbances were detected using the Landtreindr algorithm, I also excluded plots from the training dataset where there were significant deviations between field-derived dominant height and lidar-estimated canopy height. Comparisons between lidar-derived heights (95th percentile height of returns) were made with FVS-modeled dominant tree heights. Plots were discarded where lidar-derived heights were more than 100ft taller than FVS-estimated heights, where lidar-derived heights more than 50 ft shorter than FVS-estimated heights, and where lidar-derived heights were less than half of FVS-estimated heights.

⁷ Lidar attributes such as “surface volume” and “potential volume” are described in the FUSION Lidar Toolkit Manual (McGaughey, 2022).

Table 2.4. Time lags between field observations and remote sensing

FIELD YEAR	LIDAR YEARS									SATELLITE YEARS			
	Unique Plots	2012	2013	2014	2015	2016	2017	2018	Total	Unique Plots	2019	2020	Total
2010	14	3	--	2	9	--	--	--	14	355	355	0	355
2011	4	1	--	2	1	--	--	--	4	288	270	272	542
2013	528	--	524	3	8	12	24	3	574	726	693	695	1,388
2014	351	--	--	244	37	77	32	--	390	411	397	390	787
2015	433	--	--	44	262	118	28	5	457	572	560	557	1,117
2016	528	--	--	--	--	500	35	25	560	756	748	735	1,483
2017	420	--	--	--	--	--	445	--	445	512	505	500	1,005
2018	272	--	--	--	--	--	--	286	286	593	585	581	1,166
Total	2,550	4	524	295	317	707	564	319	2,730	4,199	4,099	3,717	7,816

Notes: Individual field plots may appear in more than one year of lidar or satellite data. During model fitting and cross-validation, all observations of a given plot were kept in the same cross-validation fold (i.e., used only for training or testing).

Forest structure attributes predicted

I developed models for predicting structural forest attributes as a regression task. Regression models were utilized to predict plot-level forest attributes of percent canopy cover, dominant tree height, quadratic mean diameter, live tree basal area, stem density of growing stock trees (diameter at breast height greater than or equal to five inches), and aboveground live tree carbon stocking. These plot-level attributes were generated by growing field-measured plot data forward on an annual timestep using the regionally-appropriate variant of the Forest Vegetation Simulator. I fitted models using lidar-derived features or using satellite-derived features. Results are not presented from any models trained using both lidar and satellite data as input features. Although several of the predictor variables are constrained (e.g., with canopy cover in the 0-100 range, and for all other variables to be non-negative), I did not transform (e.g., expit/inverse logit or log transformations) the target variables, instead fitting regression models directly to the raw values of these variables.

Domain shift and adaptation

The field observations collected by the various natural resource agencies are targeted and follow differing sampling protocols. These inventory campaigns were designed to build correlative models, and not necessarily to offer an unbiased sample of the forest conditions across all ownerships in the regions. As such, the domain from which training data are gathered, both in terms of predictor variables and prediction targets, may be distributed substantially differently from the domains in which these models may be applied.

There is not a simple process to determine how well correlations between predictor variables (e.g., remotely sensed attributes) and field conditions hold across space and time. In the field of ML, this general concern has come to be described as a problem of “domain shift”, and a variety of methods to support “domain adaptation” and “transfer learning” have become more common and elaborated in the past few decades, particularly for neural network models trained for computer vision and natural language processing tasks.

Given the intended use of these forest structure models for predicting conditions across all forests in the Pacific Northwestern US states, I turned to the USFS Forest Inventory & Analysis (FIA) dataset to generate estimates of the forest conditions across which these predictive models

will be applied. By design, the FIA program employs a systematic random sample of US forestlands that is well-suited for producing unbiased estimates of population attributes across US states and ecoregions.

As shown in Figure 2.2, there are several target variables and ecoregions where the field-sampled forest conditions in remote sensing plots are distributed substantially differently from conditions observed across the FIA dataset. In particular, the remote sensing plots underrepresent forest conditions with sparse or open canopy cover, that are relatively short, and that carry relatively low biomass. This type of bias between the landscape and plots sampled may correspond to reduced sampling in forests following recent timber harvests or other major disturbances, for example.

As one potential option to try and mitigate the potential for biased model training, I explored the use of weighted samples during model training. I am unaware of any other studies applying this sample weighting approach to address domain adaptation despite a substantial body of research investigating the potential for transferring correlative models between remote sensing and forest attributes across space and time. The closest approaches to correct model biases involved the fitting of semi-parametric models that add a random effect for each region to predictions made by a non-parametric model such as RF (Mauro et al., 2021) or using a linear bias-correction based on the comparison of pixel-level predictions aggregated across large regions with FIA population estimates at county and state scales (Hudak et al., 2020).

In this study, I implemented a sample weighting scheme based on nearest neighbors importance weighting (Loog, 2012). In brief, the weight assigned to each sample in the training (remote sensing plot) dataset was based on the number of times that sample is returned as one of the top 5 nearest neighbors for a plot observed in the reference (FIA) dataset. I performed this nearest neighbor search by calculating the distance between FIA and training samples using the Mahalanobis distance considering all six dimensions of forest structure being predicted (canopy cover, dominant height, basal area, growing stock stem density, and aboveground live carbon). The use of sample weighting led to (weighted) distributions of training data more closely resembling the multivariate distributions of the reference data (see Figure 2.3 and Figure 2.4).

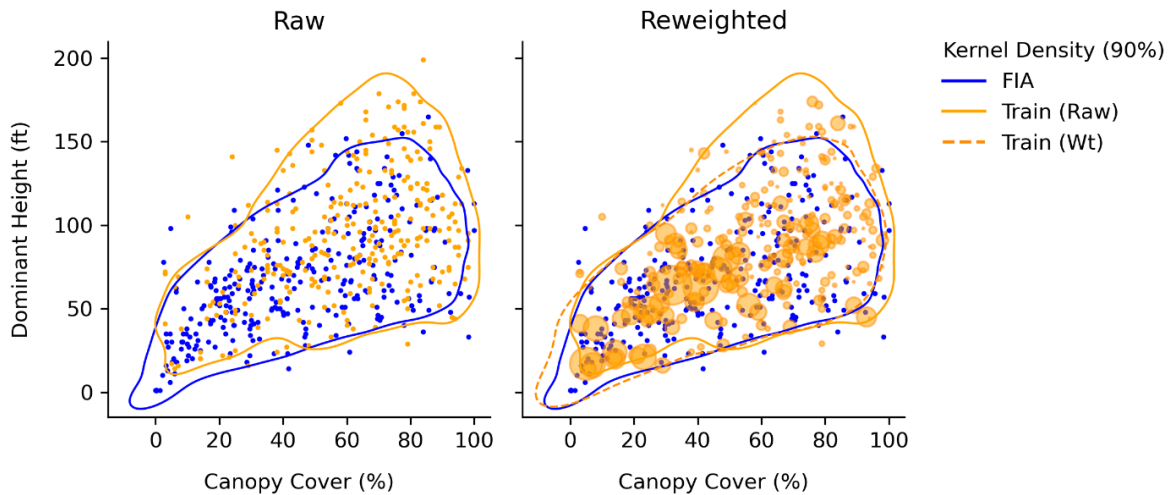


Figure 2.3. Illustration of sample re-weighting for domain adaptation

The graphs above illustrate the distributions of training data and FIA conditions across two of the six dimensions considered in this study (canopy cover and dominant height in this case). The graphs show a two-dimensional kernel density estimate with the contour line containing 90% of probability density for each dataset. Samples in the training data that are in areas underrepresented relative to the FIA data distribution are given higher weight, while those that are relatively oversampled receive less weight.

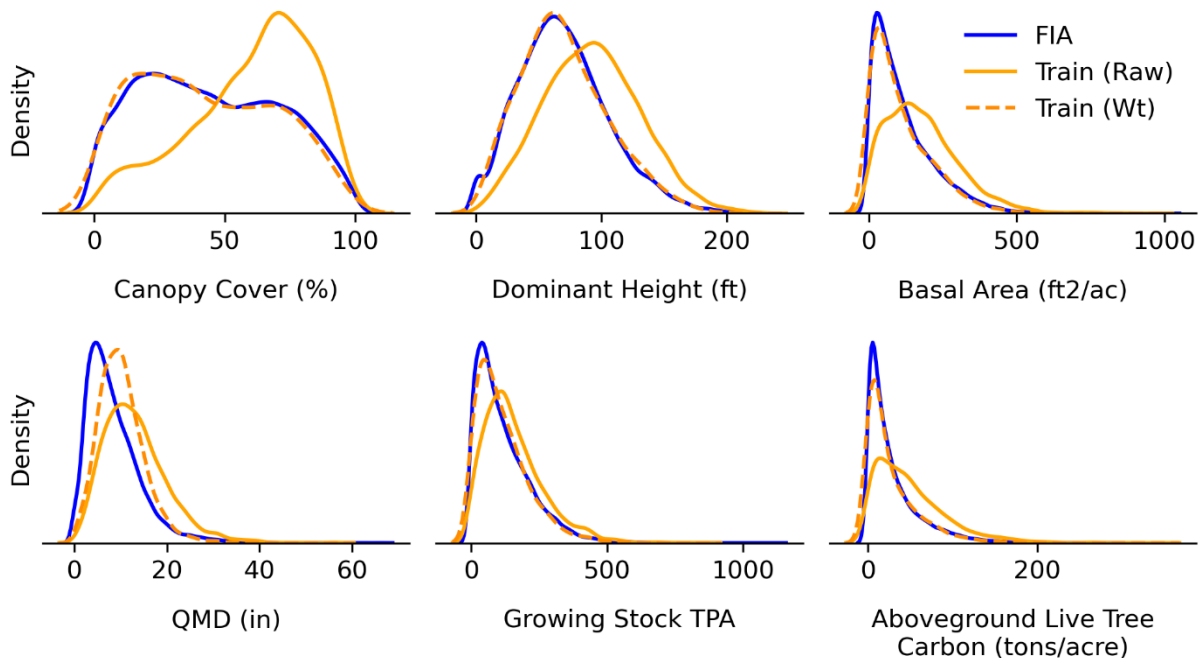


Figure 2.4. Matching FIA distributions with sample re-weighting

Different probability densities of the six forest structure attributes predicted under the raw training dataset and from FIA conditions across Oregon and Washington. By assigning different sample weights, a closer approximation to the FIA distribution (which is believed to offer an unbiased representation of the population to which the models will be applied) is achieved. Weighting was determined separately for each ecoregion.

For most of the results and findings discussed in this study, I estimated sample weights independently in each ecoregion, assigning sample weights for the Blue Mountains ecoregion based on the nearness of remote sensing forest structure samples to those observed from FIA plots within the Blue Mountains.

Development of predictive models

In this study, I fitted several regression algorithms, including: (i) linear models fit using either ElasticNet (Zou and Hastie, 2005) or Lasso regularization; (ii) k-Nearest Neighbors (kNN); (iii) Random Forests (RF) (Breiman, 2001); and (iv) Gradient Boosting Machines (GBM). All models were implemented using the scikit-learn Python package (Pedregosa et al., 2011).

I adopted a “chained” regression modeling approach. Remote sensing features (and geographic attributes) are used as the inputs to the model, which first predicts canopy cover. The input features as well as the cover-prediction are then passed as inputs to the next regressor in the chain, which predicts dominant tree height, and so on, continuing forward down the chain of regression models to predict quadratic mean diameter (QMD), etc. By the time the model is predicting its final target (aboveground live tree carbon), it also has the predicted values of canopy cover, dominant height, and diameter, basal area, QMD, and stem density available as input features along with the remotely-sensed inputs. The chained approach allows the models to exploit correlations between target variables (e.g., height, diameter, and stem density) while also offering the convenience of a single model interface which simultaneously yields predictions of multiple forest attribute given a single array of remotely-sensed inputs.

Nested cross-validation for hyperparameter tuning and model comparisons

In order to draw meaningful conclusions regarding the relative performance of alternative model types, it is critical that each model be permitted to put its best foot forward. This means incorporating a process for tuning model hyperparameters and not simply using default settings.

5-Fold Cross-Validation to Estimate Model Performance

For each round, 20% of the dataset is set aside for evaluating model performance (SCORE). The best-fitting model hyper-parameters are identified in the nested tuning loop (right). Using those best-fitting parameters, a model is fit using the 80% TRAIN subset and scored on the 20% SCORE subset that it has never seen before.

1.	SCORE	TRAIN	TRAIN	TRAIN	TRAIN
2.	...	SCORE	...	...	...
3.	...	...	SCORE	...	...
4.	...	...	...	SCORE	...
5.	...	...	...	...	SCORE

Nested 3-Fold Cross-Validation to Find Best Model Parameters

Using the 80% TRAIN subset, best-fitting model hyper-parameters are searched for three separate times using different subsets of the training data. The parameter combination which gives the highest performance across three test sets is then selected for scoring.

→	1.1	TEST	TUNE	TUNE
	1.2	...	TEST	...
	1.3	...	...	TEST

This nested tuning loop is repeated in each of the five scoring rounds

For a particular type of model (e.g., a Random Forest), the nested cross-validation approach generates five scores (one for each round of cross-validation). The distribution of these scores help indicate how a well-tuned model performs on data it hasn't seen before. When that model is ultimately applied in production, it is fitted using all available data.

Figure 2.5. Illustration of nested cross validation for model tuning and validation.

To tune hyperparameters of each algorithm, I employed grid search cross-validation. For the linear models (ElasticNet and Lasso), I tuned the “alpha” parameter which modifies how strongly the model penalizes the inclusion of additional terms, and in the case of ElasticNet, I also tuned the “l1_ratio” parameter, which controls how close to pure L1 or pure L2 penalty is applied⁸. I tuned three parameters in the kNN algorithm, including the number of neighbors, whether the k-nearest neighbors are weighted when coming up with a regression prediction, and whether the distance metric uses Minkowski or Manhattan distance. For the RF algorithm, I tuned the “n_estimators” (number of trees), “max_features” (predictor variables used by each tree), “max_depth” (of each tree), and “max_samples” (proportion of training samples used in building each tree). For the GBM algorithm, I tuned the “max_iter” (number of trees), “min_samples_leaf” (number of samples for each leaf in each tree), “max_depth” (per tree), and “learning_rate”, which controls or regularizes the scale of updates the GBM algorithm uses to build and refine decision trees.

In general, I adopted these algorithms and model-tuning workflows because I was not primarily interested in the specific settings for these hyperparameters, but rather approached this problem with the expectation that all these algorithms should be tuned before having their performance evaluated and compared. These tuning processes, when paired with nested cross-validation, provide a valuable opportunity to limit the potential for overfitting or excessive computation (without performance improvements) that can easily occur with algorithms like RF and GBM when large numbers of trees and very deep trees are used.

Regional hold-out methods to enable assessments of model generalization and transferability across ecoregions

To investigate inter-regional model transferability, I adopted and compared four cross-validation strategies:

- “Visitor” models – trained using observations from a single region and evaluated on data from a different region

⁸ The alpha parameters in this models as implemented in scikit-learn is equivalent to the lambda parameter as implemented in the R package glmnet. The l1_ratio parameter in scikit-learn is equivalent to the alpha parameter in glmnet.

- “Outsider” models – trained using observations from all regions *except* the region used for evaluation, but without using region as an explicit feature
- “Insider” models – trained using observations from a single region and evaluated on held-out data from that region
- “Global” models – trained using observations from all regions and evaluated on held-out data in each region

By fitting and evaluating models using strategically isolated training and testing datasets, it was possible to more meaningfully characterize how much information and generalization capability is gained by incorporating observations from new regions that may have substantially different environmental conditions and forest types from a region in which these models might be applied (extrapolated). I was particularly interested in characterizing any changes in model performance given a choice between applying Insider models and Global models. For each type of model, I quantified how the average performance in each ecoregion changes for tuned Global versus Insider models changed using several indicators.

Performance indicators and model comparison

To refer to the various models compared, I adopted a naming convention combining the algorithm used (e.g., RF, KNN), geographic scope of data (e.g., INSIDER, GLOBAL), and source of data (e.g., LIDAR, SATELLITE). For example, GBM-GLOBAL-LIDAR refers to a model employing the Gradient Boosting Machines algorithm trained using data from all ecoregions with remotely-sensed features derived from lidar, whereas RF-INSIDER-SATELLITE refers to a model employing the Random Forests algorithm trained and tested with data from a single ecoregion with remotely-sensed features derived from satellite imagery.

To evaluate models, I followed the recommendations of Riemann et al. (2010) in generating multiple performance metrics including root mean squared error (RMSE), mean absolute error (MAE), mean error (bias), and normalized versions of RMSE, MAE, and bias to quantify predictive errors as a percentage of the mean value for a particular statistic (i.e., NRMSE for RMSE and MAPE for MAE, and relative bias). I also employed the Kolmogorov-Smirnov statistic as a distance measure (KS distance) to quantify inter-regional differences in distributions of individual forest structure target variables (Riemann et al., 2010).

Model performance scores were calculated based on nested cross-validation with five outer folds in each of nine ecoregions. The cross-validation scores in each ecoregion were treated as distinct datasets, and model comparison was conducted following (Demšar, 2006) as implemented and visualized by the Autorank package in Python (Herbold, 2020). The performance of two models was considered significantly different if the corresponding average ranks differ by at least the Critical Difference (CD) determined by the Nemenyi test (Nemenyi, 1963).

To characterize the potential for model transferability, following Fekety et al. (2015) I considered Visitor and Global models to offer comparable performance to Insider models in any given ecoregion if the average performance metric of the Visitor or Global model was better than or equal to the mean plus one standard deviation for that performance metric by the Insider model, which was estimated through five-fold cross-validation.

Reproducibility

Workflows used to train and validate models and to summarize and visualize data and model performance and comparisons were implemented through a series of Jupyter Notebooks. The training data (summarized to tabular form) along with the code and Notebooks necessary to reproduce these analyses are published on GitHub with a permissible open-source license: <https://github.com/Ecotrust/ForestMapping>. The raw training data (e.g., point clouds, satellite images, etc.) have not yet been published.

2.3 RESULTS

Overall predictive performance

Lidar-based models consistently outperformed satellite-based models for every algorithm and forest structure attribute predicted in terms of model precision (see Figure 2.6 and Figure 2.7). For all six target variables, the ranks of the top-performing lidar variables were always found to be statistically significantly different (higher) than the best-performing satellite models. For canopy cover, lidar-based models consistently outperformed their satellite-based counterparts, but the differences were more muted, with several satellite-based models showing performance that was not statistically significantly different from mid- and lower-ranking lidar models.

For every target variable, either the GBM-GLOBAL-LIDAR or RF-GLOBAL-LIDAR model was the top-ranked model. The corresponding Insider versions of these models (GBM-INSIDER-LIDAR and RF-INSIDER-LIDAR) were never lower than the Global models to a statistically significant degree. For linear models (both ElasticNet and Lasso), Insider models commonly outranked Global models in terms of model precision (i.e., (N)RMSE and MA(P)E).

In terms of model precision, the ordering of model performance across algorithms was similar for both lidar- and satellite-based models, with GBM generally showing top performance, followed by RF, then the linear models (Lasso and ElasticNet), with kNN often showing the worst performance. In the case of QMD predictions, lidar-based kNN models showed worse performance on average than satellite-based linear Insider models.

Models trained with data from the regions in which they were tested (Global and Insider models) generally showed predictive performance for lidar-based models in the range of 25% NRMSE for canopy cover, 15-20% for dominant height, 30% for basal area, 30-35% for QMD, 40% for growing stock stem density, and 30-35% for aboveground live tree carbon (see Figure 2.8 and Figure 2.9). For satellite models, Insider and Global models showed slightly worse performance, with NRMSE generally in the range of 35% for canopy cover, 30% for dominant height, 40% for basal area, 35-40% for QMD, 50-60% for growing stock stem density, and 40% for aboveground live tree carbon (see Figure 2.8 and Figure 2.9).

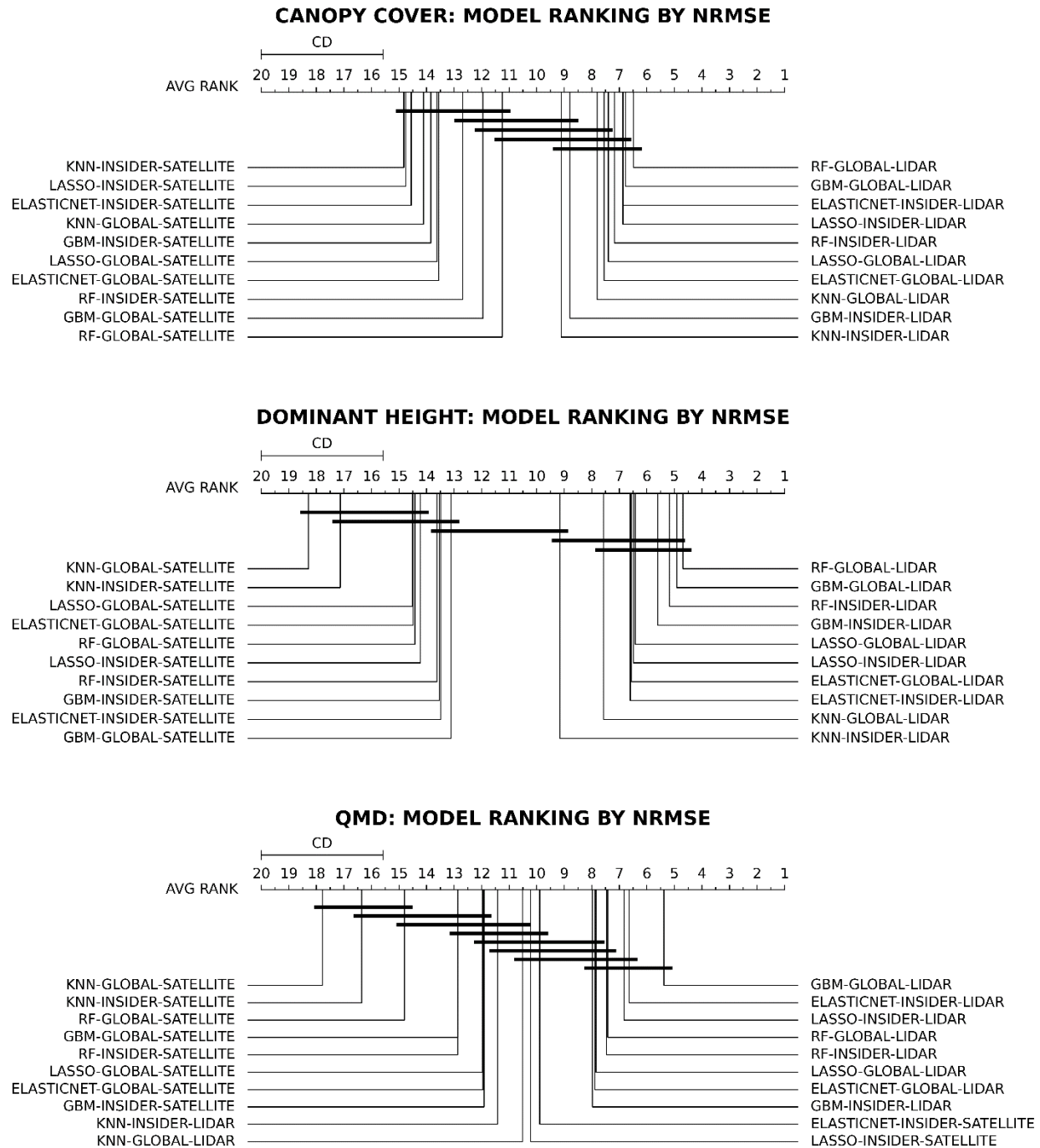


Figure 2.6. Rankings of forest structure predictive models based on NRMSE (part 1 of 2). Each graph displays the average ranking of models based on weighted training data for each combination of algorithm (GBM, ElasticNet, Lasso, RF, and kNN), scope (insider or global) and data source (lidar or satellite) across all ecoregions following Demšar (2006). Models that are not statistically significantly different from each other are displayed as crossing the same horizontal black bar below the rank scale.

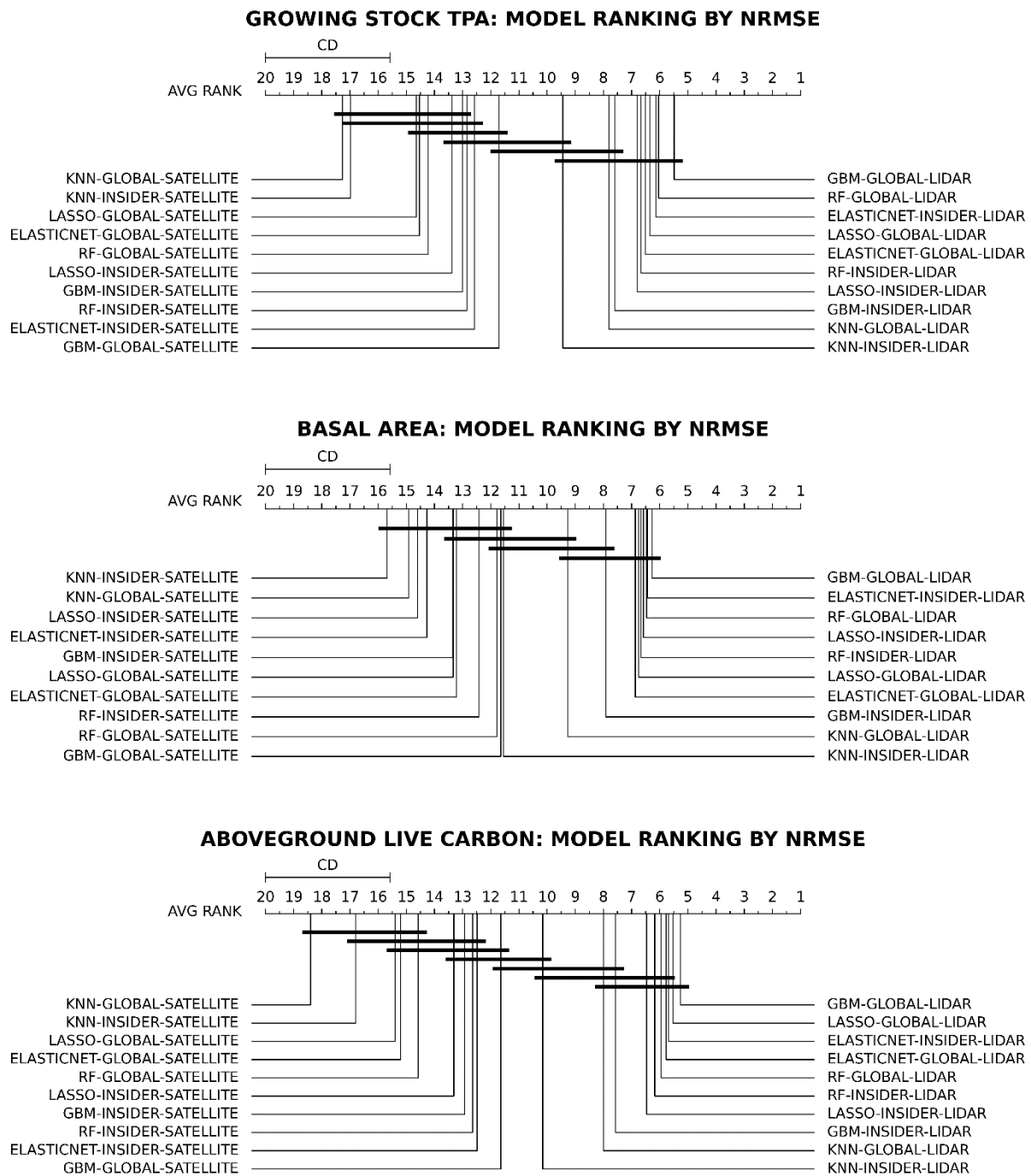


Figure 2.7. Rankings of forest structure predictive models based on NRMSE (part 2 of 2). Each graph displays the average ranking of models based on weighted training data for each combination of algorithm (GBM, ElasticNet, Lasso, RF, and kNN), scope (insider or global) and source (lidar or satellite) across all ecoregions following Demšar (2006). Models that are not statistically significantly different from each other are displayed as crossing the same horizontal black bar below the rank scale.

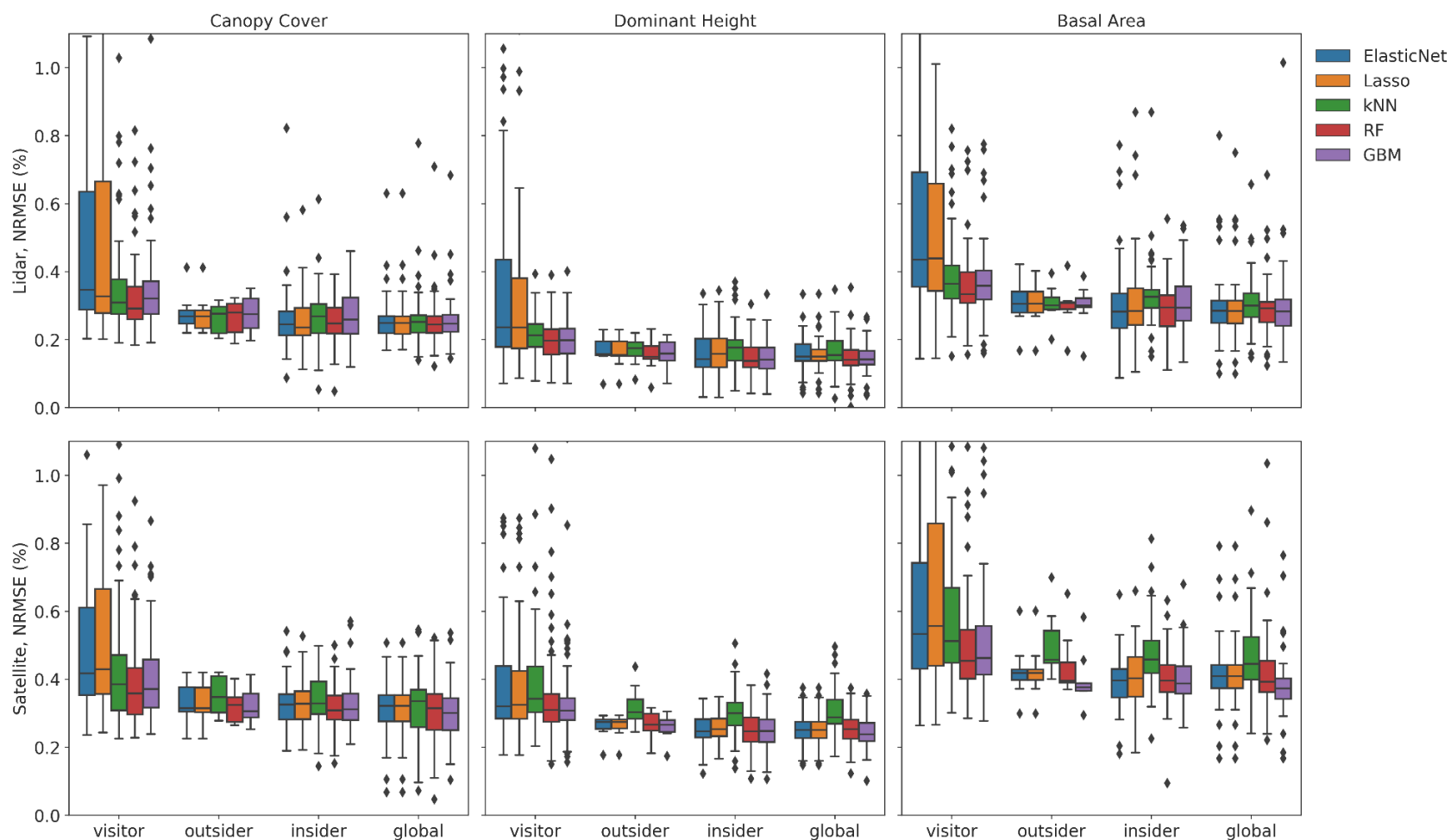


Figure 2.8. Predictive performance (Normalized Root Mean Square Error, NRMSE) of lidar- and satellite-based models on three forest structure attributes (1 of 2).

The middle line in each box displays the median, the upper and lower edges the 25th and 75th percentiles, and whiskers for 5th and 95th percentiles. Individual observations outside these ranges are shown as points. For Global and Insider models, each box summarizes five cross-validation scores across nine ecoregions (45 values); for Visitor models, each box summarizes 72 scores of algorithms trained in one of nine ecoregions applied in another (for each of nine regions, there are eight Visitor models scored once); and for Outsider models, each box summarizes nine scores (one Outsider model applied to each of nine ecoregions).

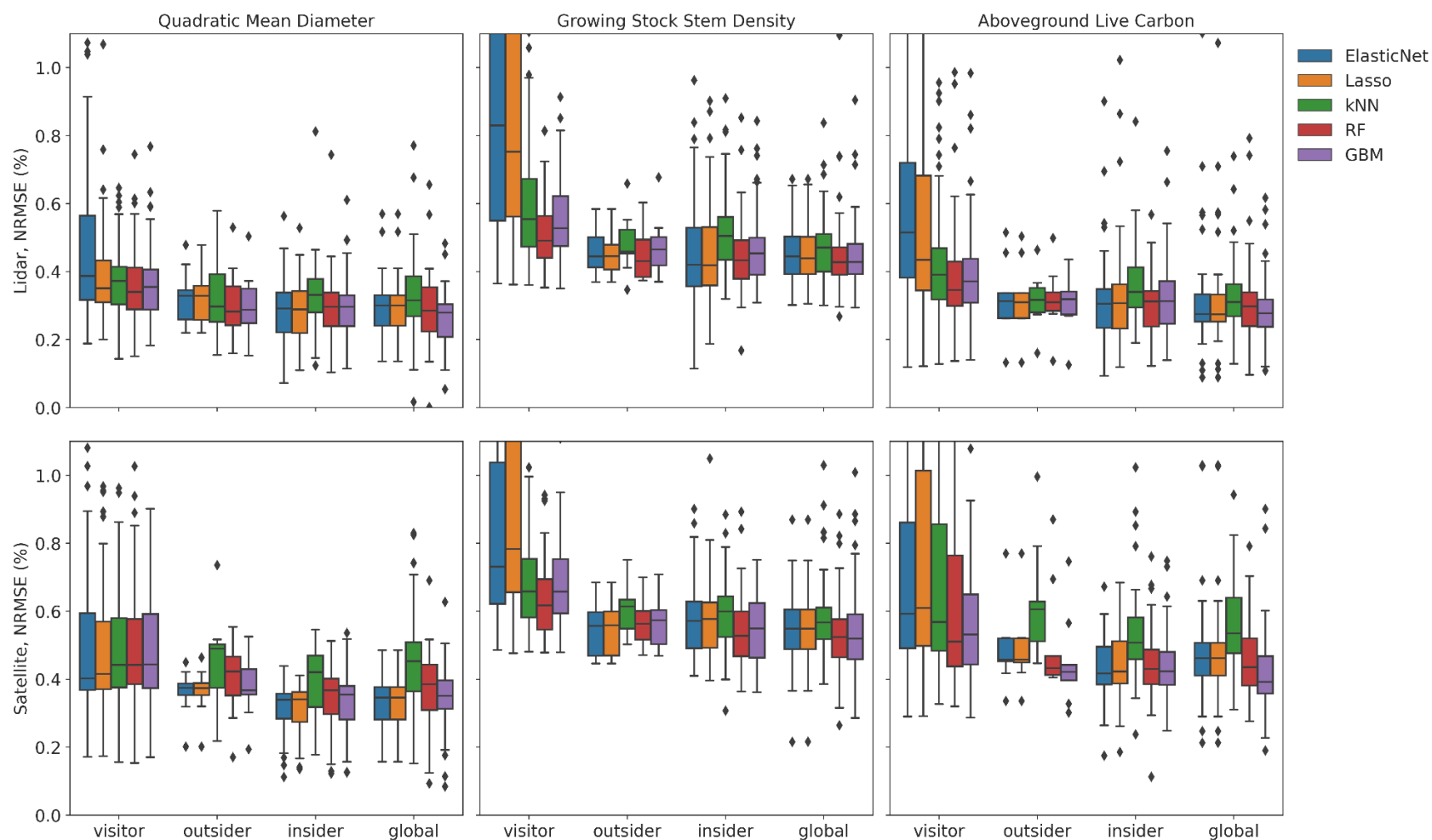


Figure 2.9. Predictive performance (Normalized Root Mean Square Error, NRMSE) of lidar- and satellite-based models on three forest structure attributes (2 of 2).

The middle line in each box displays the median, the upper and lower edges the 25th and 75th percentiles, and whiskers for 5th and 95th percentiles. Individual observations outside these ranges are shown as points. For Global and Insider models, each box summarizes five cross-validation scores across nine ecoregions (45 values); for Visitor models, each box summarizes 72 scores of algorithms trained in one of nine ecoregions applied in another (for each of nine regions, there are eight Visitor models scored once); and for Outsider models, each box summarizes nine scores (one Outsider model applied to each of nine ecoregions).

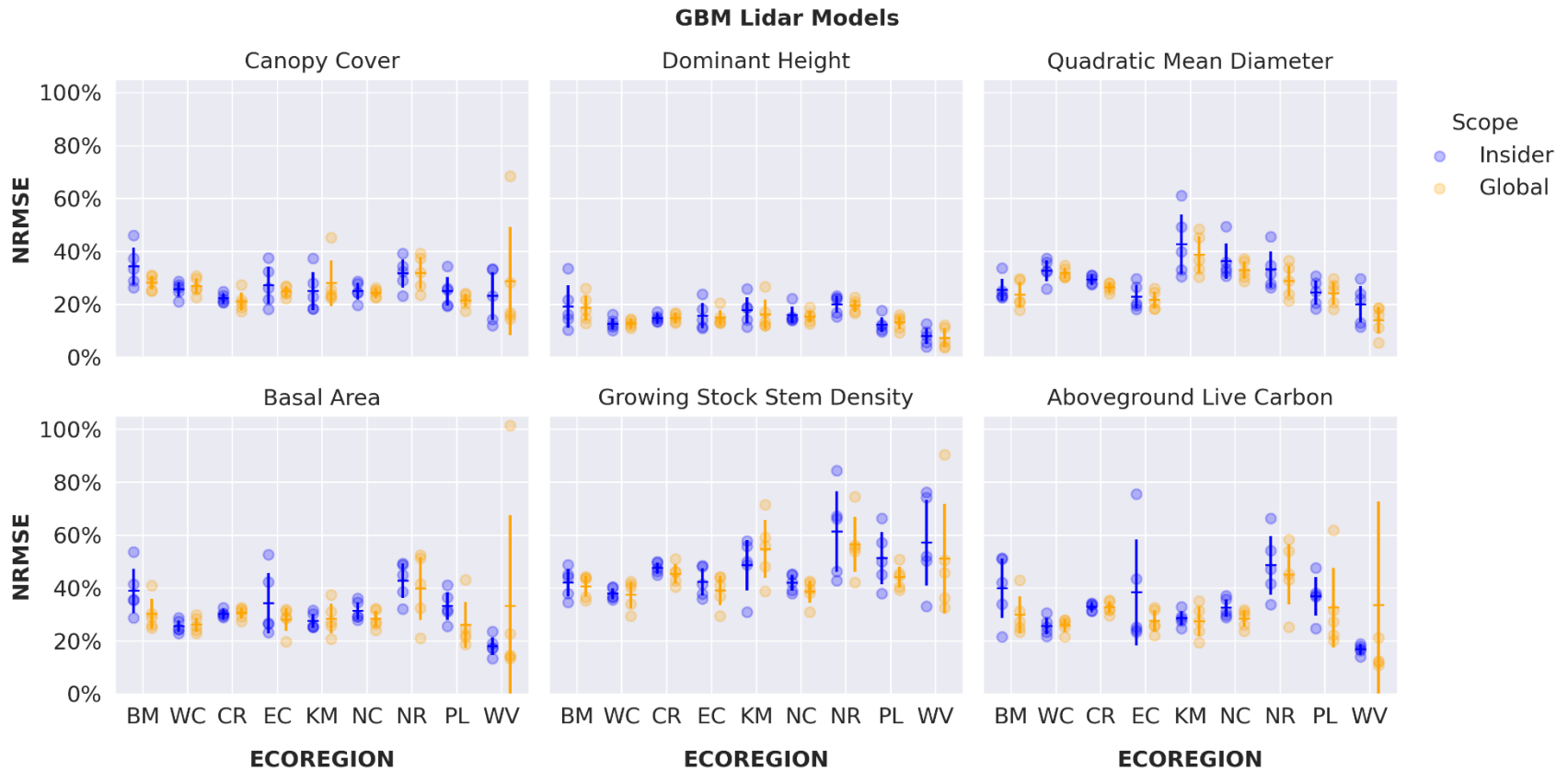


Figure 2.10. Performance (Normalized Root Mean Square Error, NRMSE) of GBM-GLOBAL and GBM-INSIDER lidar models across all ecoregions.

These graphs show the scores from five rounds of cross-validation for the GBM Insider and Global models trained from lidar data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.

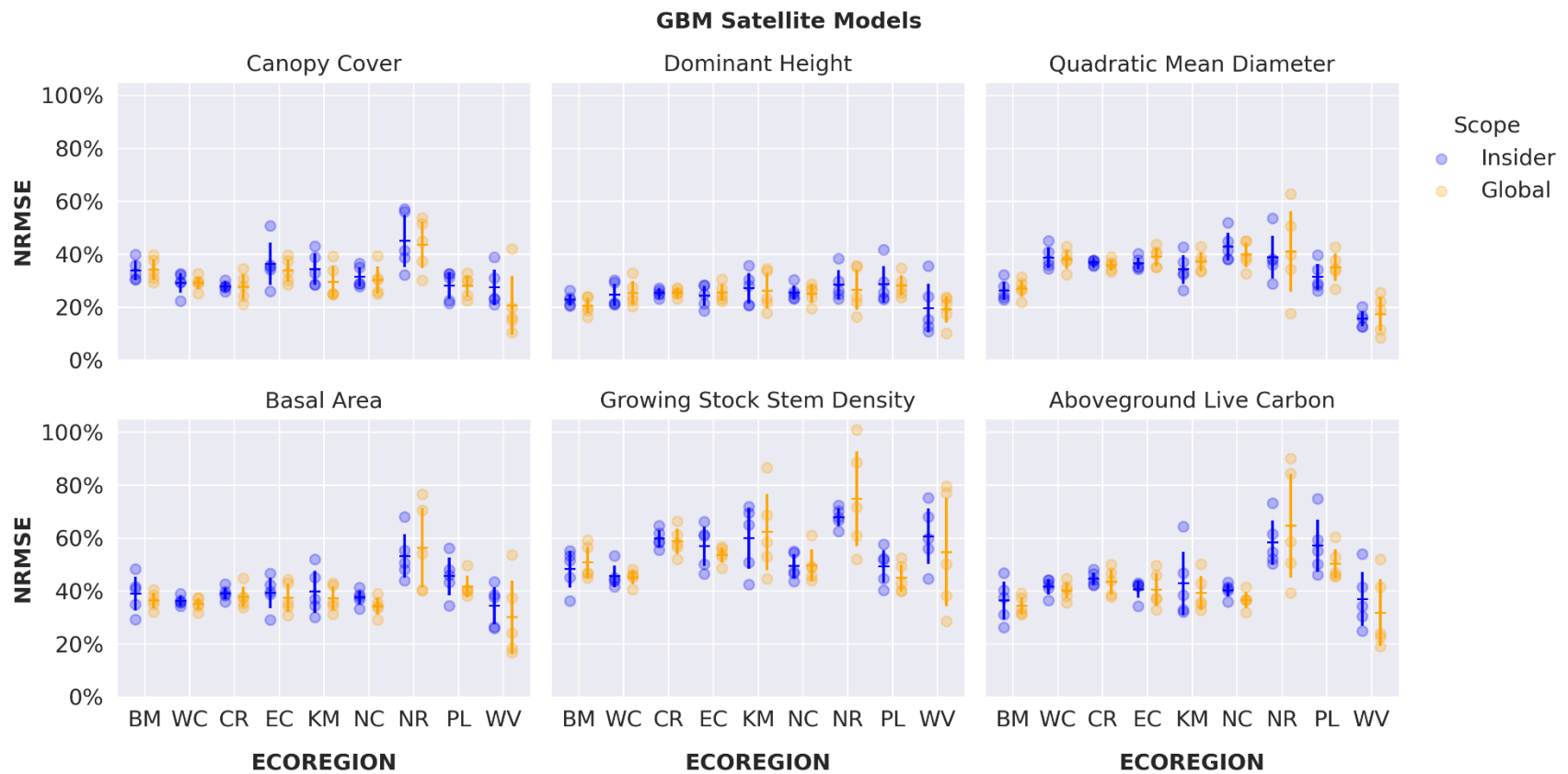


Figure 2.11. Performance (Normalized Root Mean Square Error, NRMSE) of GBM-GLOBAL and GBM-INSIDER satellite models across all ecoregions.

These graphs show the scores from five rounds of cross-validation for the GBM Insider and Global models trained from satellite data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.

Model precision (e.g., NRMSE) was generally consistent in each ecoregion, although the variation in cross-validation scores was often larger for regions that had smaller numbers of observations (see GBM model performance across regions in Figure 2.10 and Figure 2.11 and for all other algorithms in Figure S2.33 through Figure S2.40).

When model performance was considered in terms of average bias, linear models, particularly linear Insider models, more commonly showed up among the top-ranked models (see Figure 2.12 and Figure 2.13). Lidar-based models also tended to show smaller biases than their satellite-based counterparts, but this performance advantage was not as strong or clearcut as in the case of NRMSE or MAPE, with several instances of satellite-based models showing rankings for average bias that were not statistically significantly worse than the top-performing lidar models. kNN models were consistently the lowest-ranking models in terms of average bias.

Among Global and Insider models, average bias was much more highly variable among the different algorithms employed than was the case with model precision, especially for satellite-based models. Linear models, both ElasticNet and Lasso generally showed limited average bias for both lidar and satellite-based models, while ML models commonly showed negative biases and tended to under-predict field-recorded forest structure attributes. Satellite-based kNN and RF models showed pronounced negative biases on the order of -20 to -30%. Lidar-based GBM models generally showed comparable levels of bias as linear models (in the range of -5 to +5 %), but substantial negative biases were observed for GBM satellite-based models.

Across ecoregions, the performance of GBM Global and Insider models in terms of bias varied modestly (see Figure 2.16 and Figure 2.17 for bias of GBM models across regions and Figure S2.41 through Figure S2.48 for bias of all other models across regions).

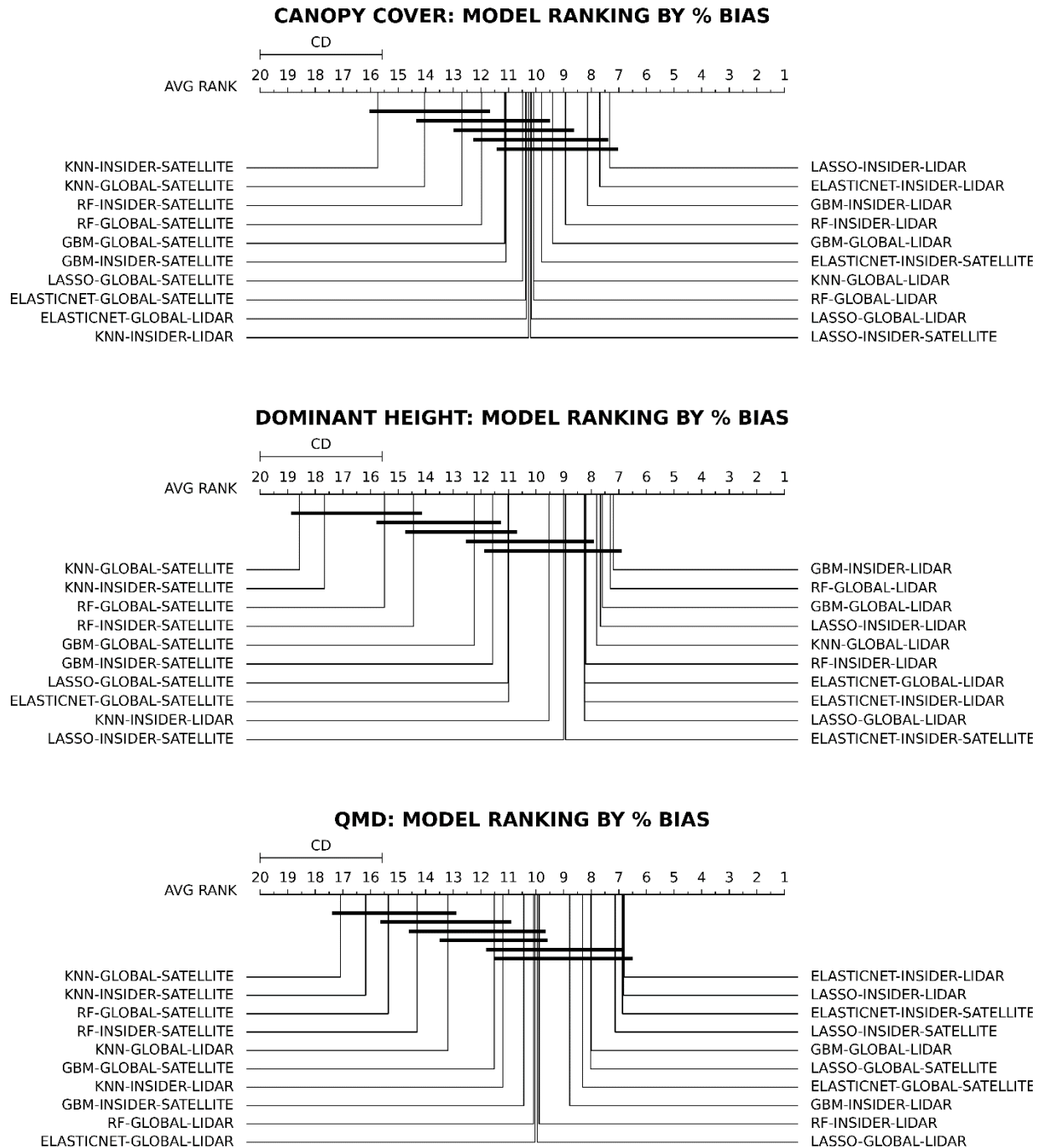


Figure 2.12. Rankings of forest structure models based on relative bias (part 1 of 2).
Each graph displays the average ranking of models based on weighted training data for each combination of algorithm, scope (insider or global) and source (lidar or satellite) across all ecoregions following Demšar (2006). Models that are not statistically significantly different from each other are displayed as crossing the same horizontal black bar below the rank scale.

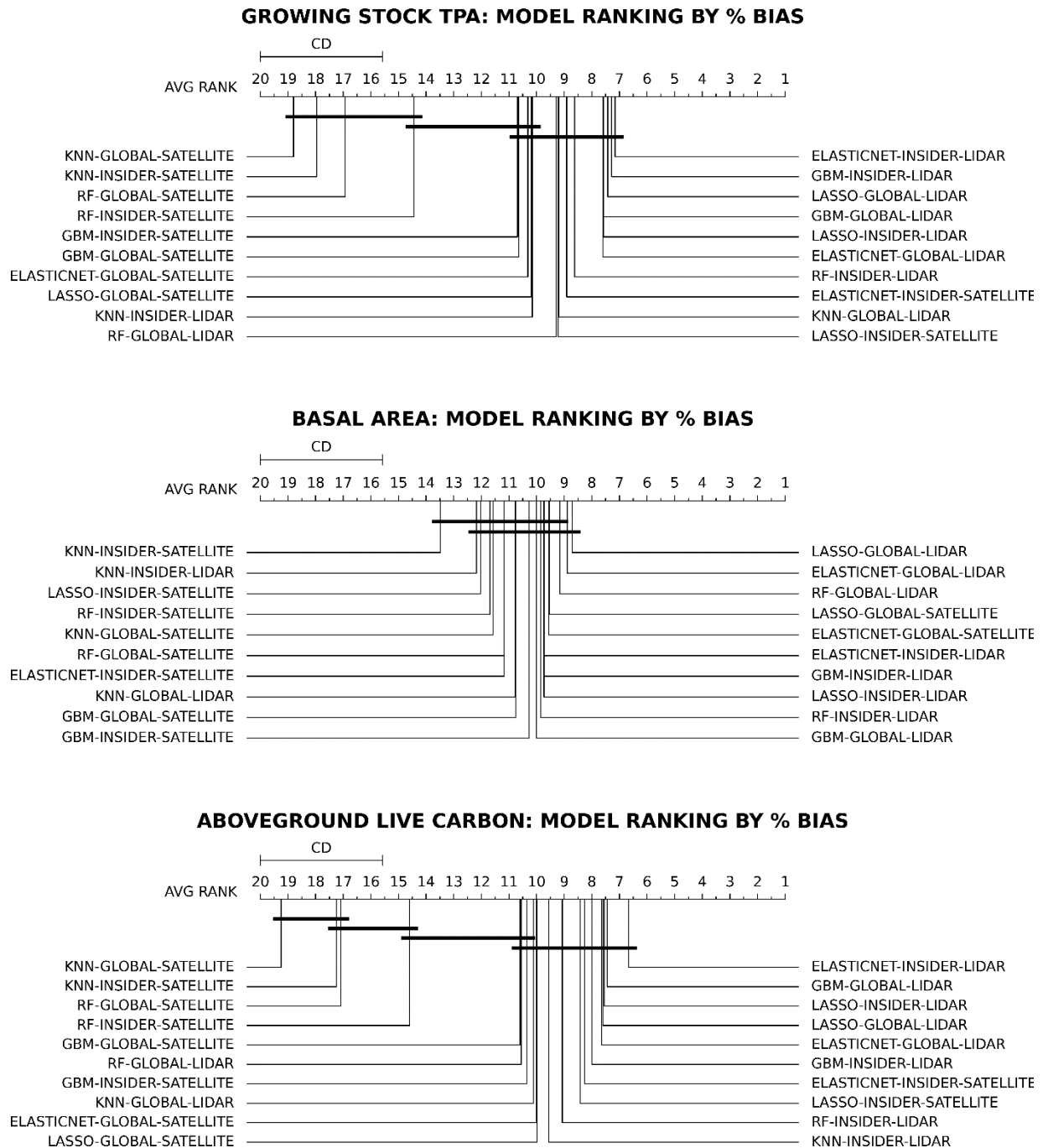


Figure 2.13. Rankings of forest structure models based on relative bias (part 2 of 2). Each graph displays the average ranking of models based on weighted training data for each combination of algorithm, scope (insider or global) and source (lidar or satellite) across all ecoregions following Demšar (2006). Models that are not statistically significantly different from each other are displayed as crossing the same horizontal black bar below the rank scale.

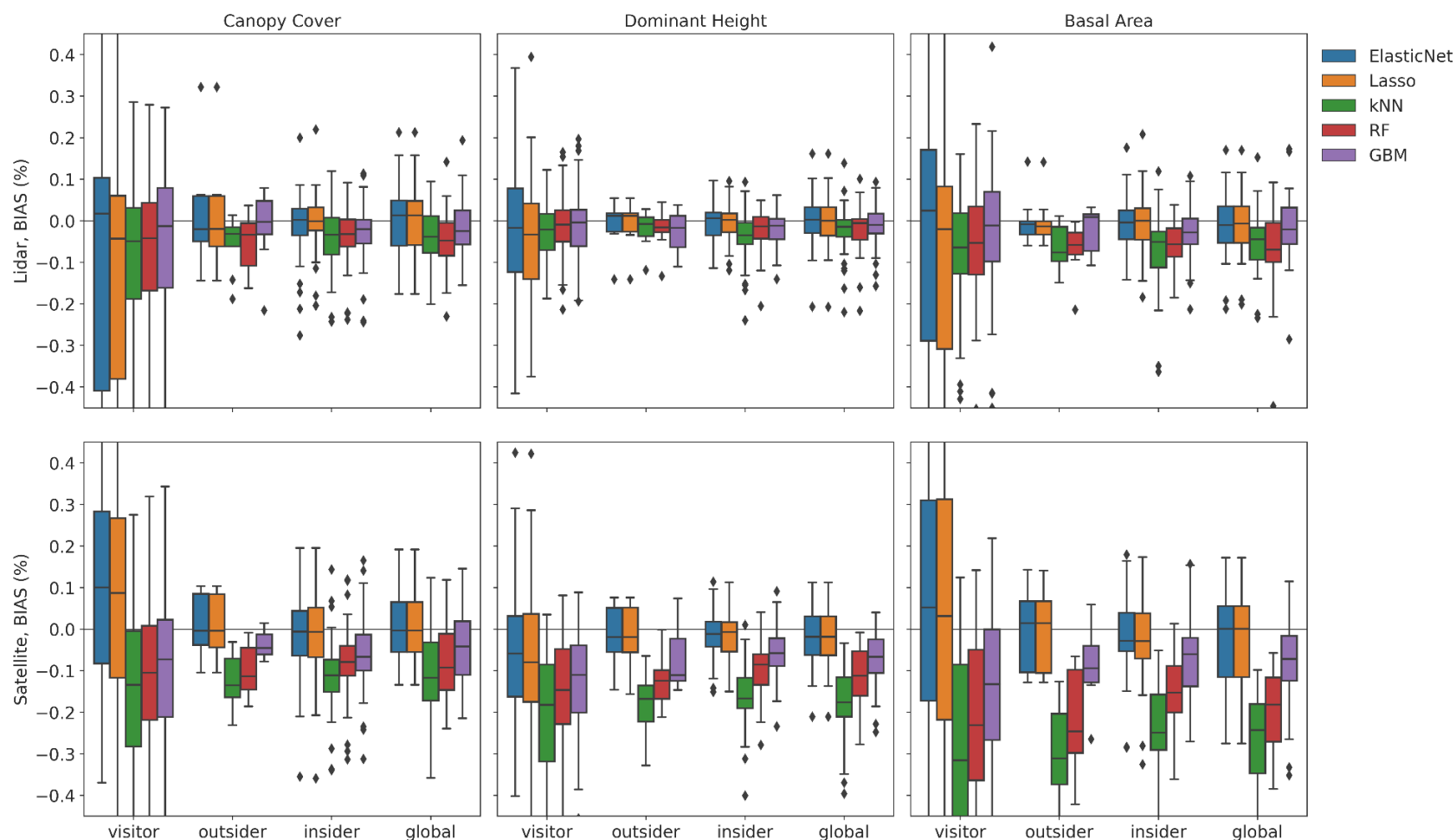


Figure 2.14. Predictive performance (% bias) of lidar- and satellite-based models on three forest structure attributes (1 of 2). *The middle line in each box displays the median, the upper and lower edges the 25th and 75th percentiles, and whiskers for 5th and 95th percentiles. Individual observations outside these ranges are shown as points. For Global and Insider models, each box summarizes five cross-validation scores across nine ecoregions (45 values); for Visitor models, each box summarizes 72 scores of algorithms trained in one of nine ecoregions applied in another (for each of nine regions, there are eight Visitor models scored once); and for Outsider models, each box summarizes nine scores (one Outsider model applied to each of nine ecoregions).*

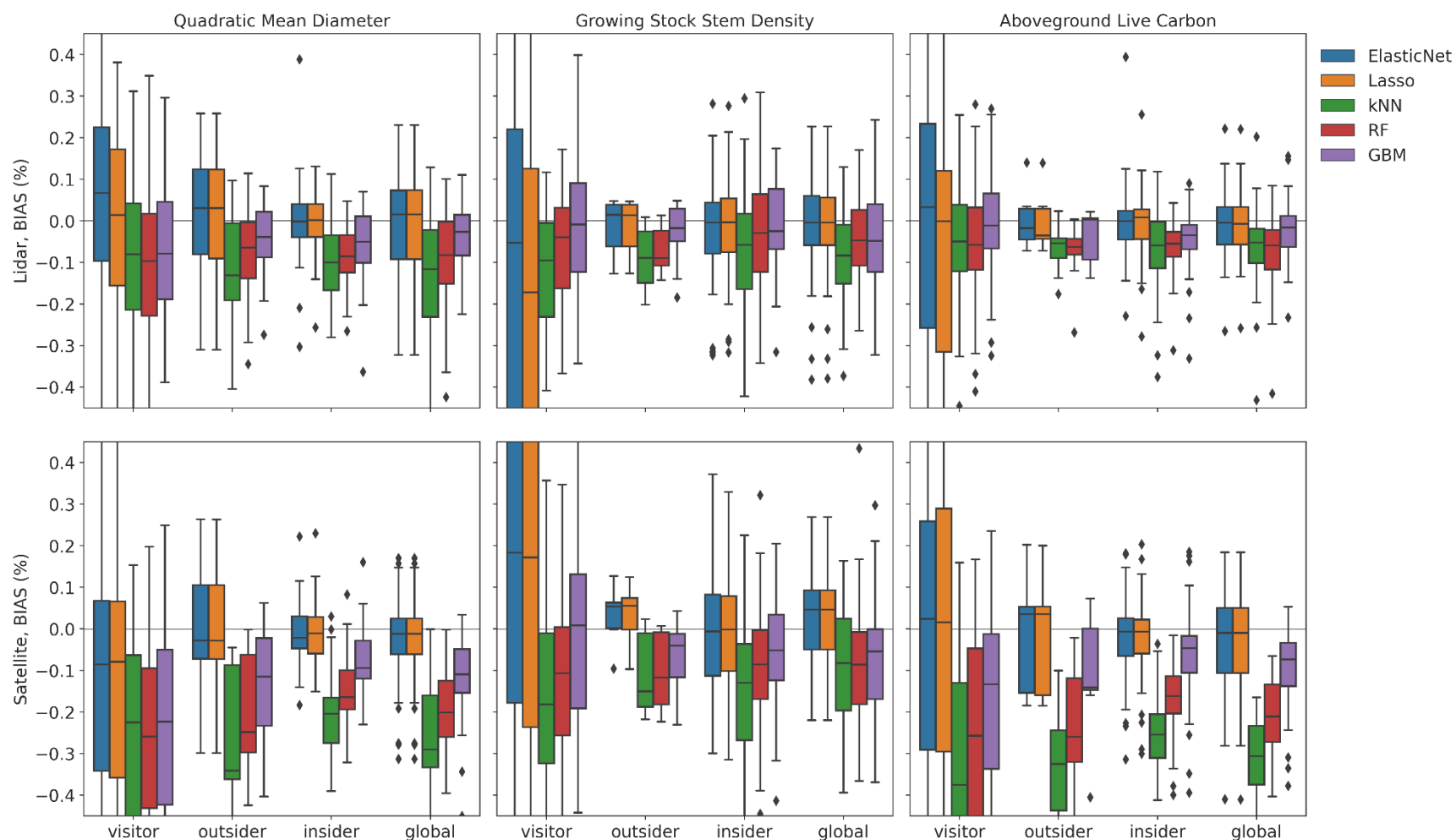


Figure 2.15. Predictive performance (% bias) of lidar- and satellite-based models on three forest structure attributes (2 of 2). The middle line in each box displays the median, the upper and lower edges the 25th and 75th percentiles, and whiskers for 5th and 95th percentiles. Individual observations outside these ranges are shown as points. For Global and Insider models, each box summarizes five cross-validation scores across nine ecoregions (45 values); for Visitor models, each box summarizes 72 scores of algorithms trained in one of nine ecoregions applied in another (for each of nine regions, there are eight Visitor models scored once); and for Outsider models, each box summarizes nine scores (one Outsider model applied to each of nine ecoregions).

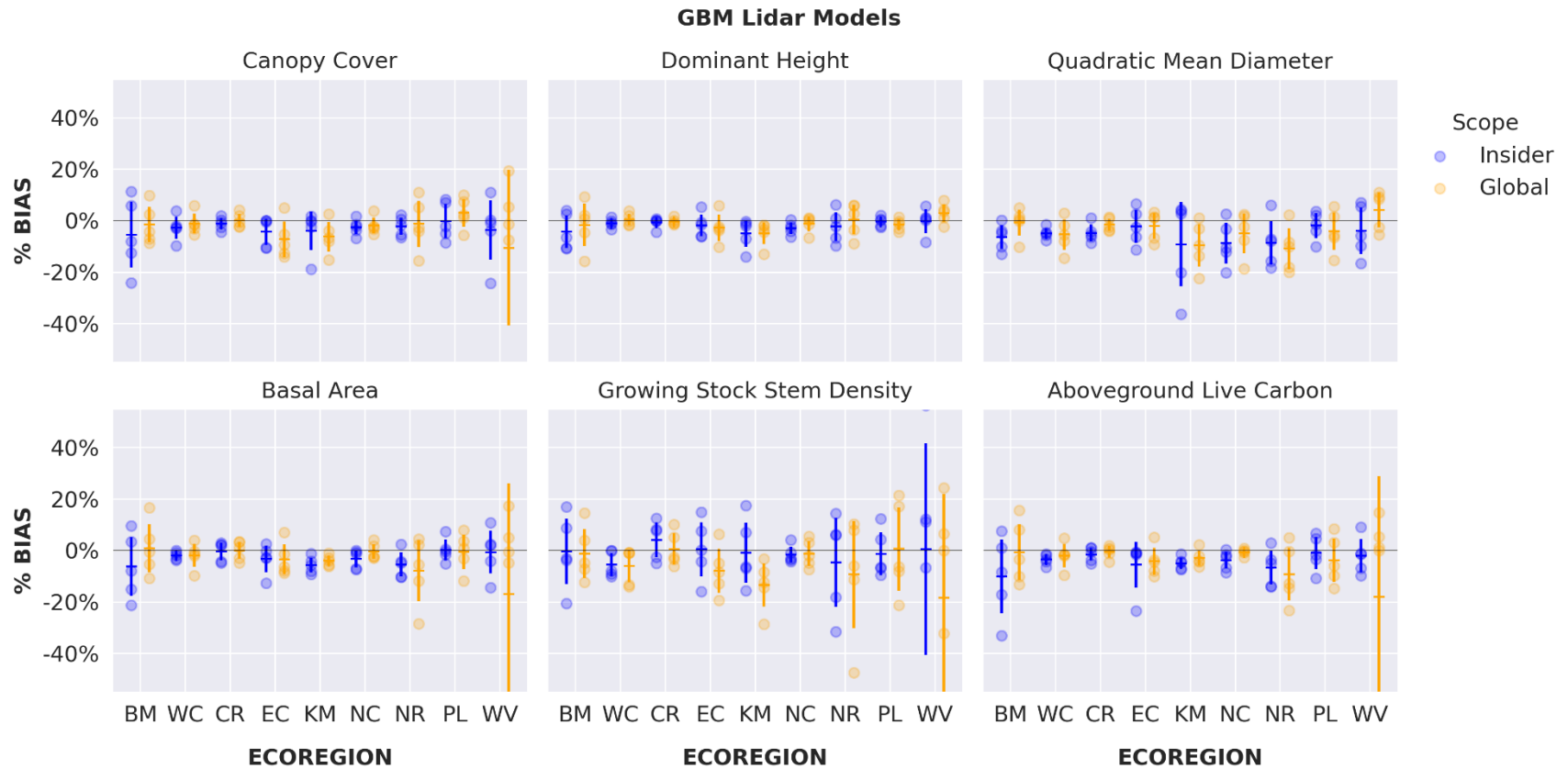


Figure 2.16. Performance (% bias) of GBM-GLOBAL and GBM-INSIDER lidar models across all ecoregions.
These graphs show the scores from five rounds of cross-validation for the GBM Insider and Global models trained from lidar data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.

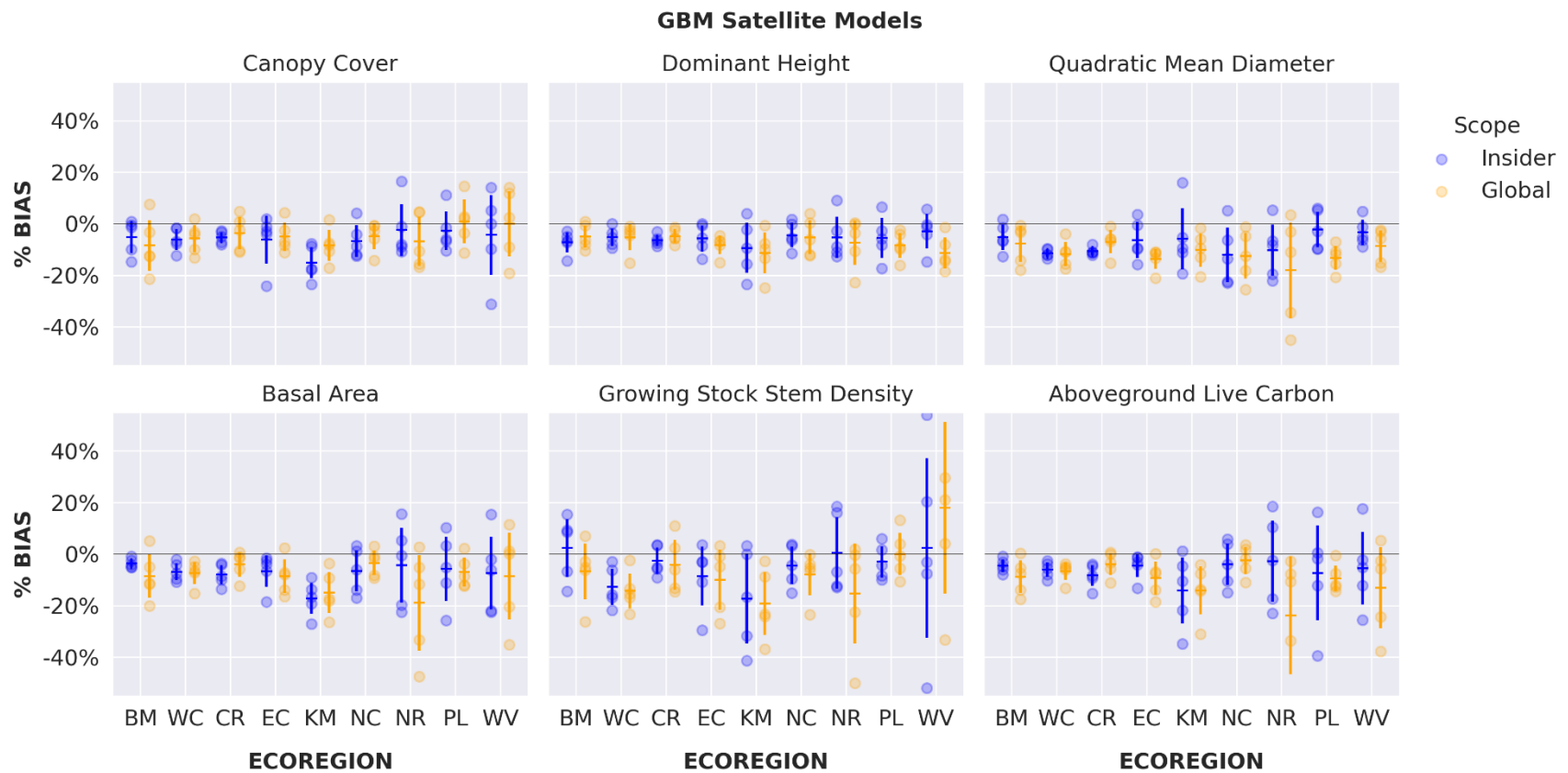


Figure 2.17. Performance (% bias) of GBM-GLOBAL and GBM-INSIDER satellite models across all ecoregions. *These graphs show the scores from five rounds of cross-validation for the GBM Insider and Global models trained from satellite data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.*

When looking across the range of each target variable, all models showed biases related to the magnitude of the target variable. Observed and predicted values for GBM (see Figure 2.18 and Figure 2.19) and Lasso models (see Figure 2.20 and Figure 2.21) illustrate these trends in model bias well. All models tended to overpredict canopy cover at low levels and to underpredict it at high levels. Lidar models showed relatively limited bias in the prediction of field-estimated dominant height, while satellite-based models more strongly overpredicted the height of short plots and underpredicted the height of taller plots. Predictions of basal area and aboveground live tree carbon by both lidar and satellite models generally showed little-to-no-bias over most of the range of these variables, but did follow a common “saturating” pattern observed in many remote sensing studies which very high levels of basal area and carbon stocking tending to be underpredicted.

The shape of prediction biases across the range of each target variable were generally the same between lidar- and satellite-based models, though satellite-based models generally showed more accentuated patterns of bias that were observed in lidar-based models (see, for example, Figure 2.18 and Figure 2.19). Similarly, linear models showed even more accentuated patterns of bias across the range of each predictor variable than ML models like GBM and RF (see Figure 2.20 and Figure 2.21). Linear models showed strong saturating effects for predictions of QMD, growing stock stem density, basal area, and aboveground live tree carbon that were particularly prominent in satellite-based linear models.

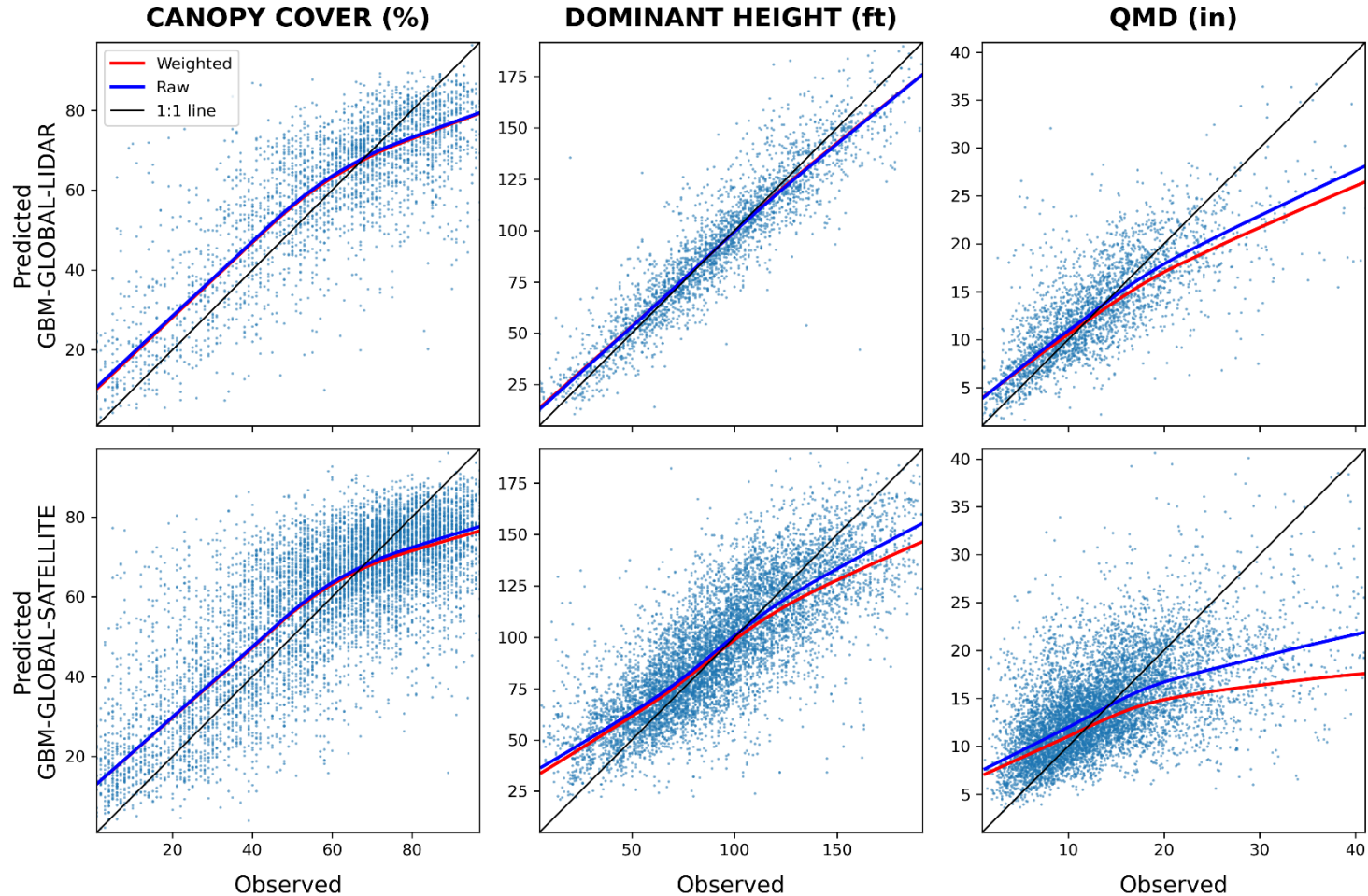


Figure 2.18. Correspondence between predictions and observations for top-performing satellite and lidar models (1 of 2).

These graphs show the predicted and observed values for top-performing lidar (top row) and satellite (bottom row) models. Values above illustrate predictions on data withheld from model training using five 80-20% folds. The blue line shows a locally weighted scatterplot smoothing function for the model trained on raw (unweighted) data, and red line for the model trained on weighted training data.

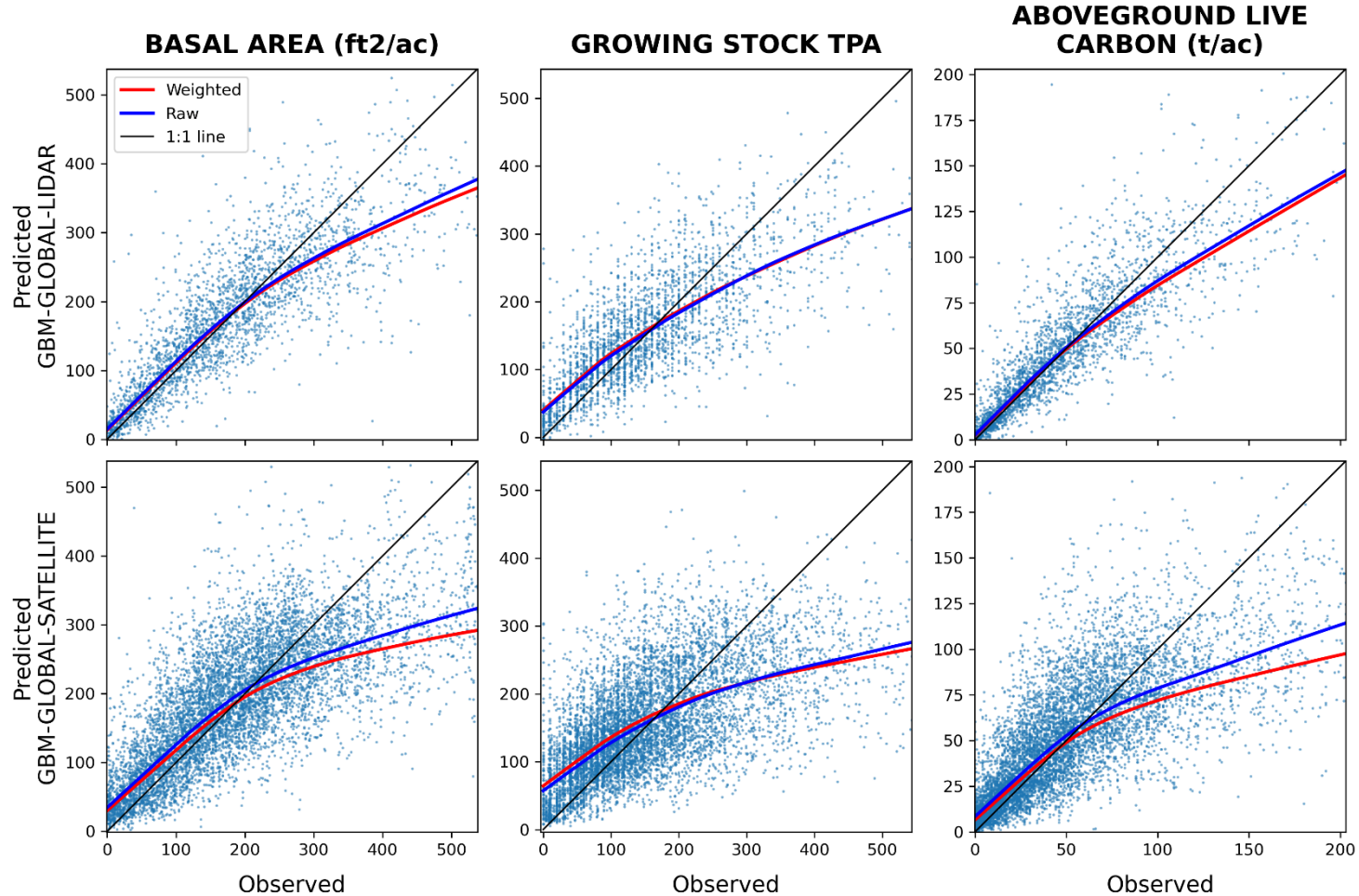


Figure 2.19. Correspondence between predictions and observations for top-performing satellite and lidar models (2 of 2). These graphs show the predicted and observed values for top-performing lidar (top row) and satellite (bottom row) models. Values above illustrate predictions on data withheld from model training using five 80-20% folds. The blue line shows a locally weighted scatterplot smoothing function for the model trained on raw (unweighted) data, and red line for the model trained on weighted training data.

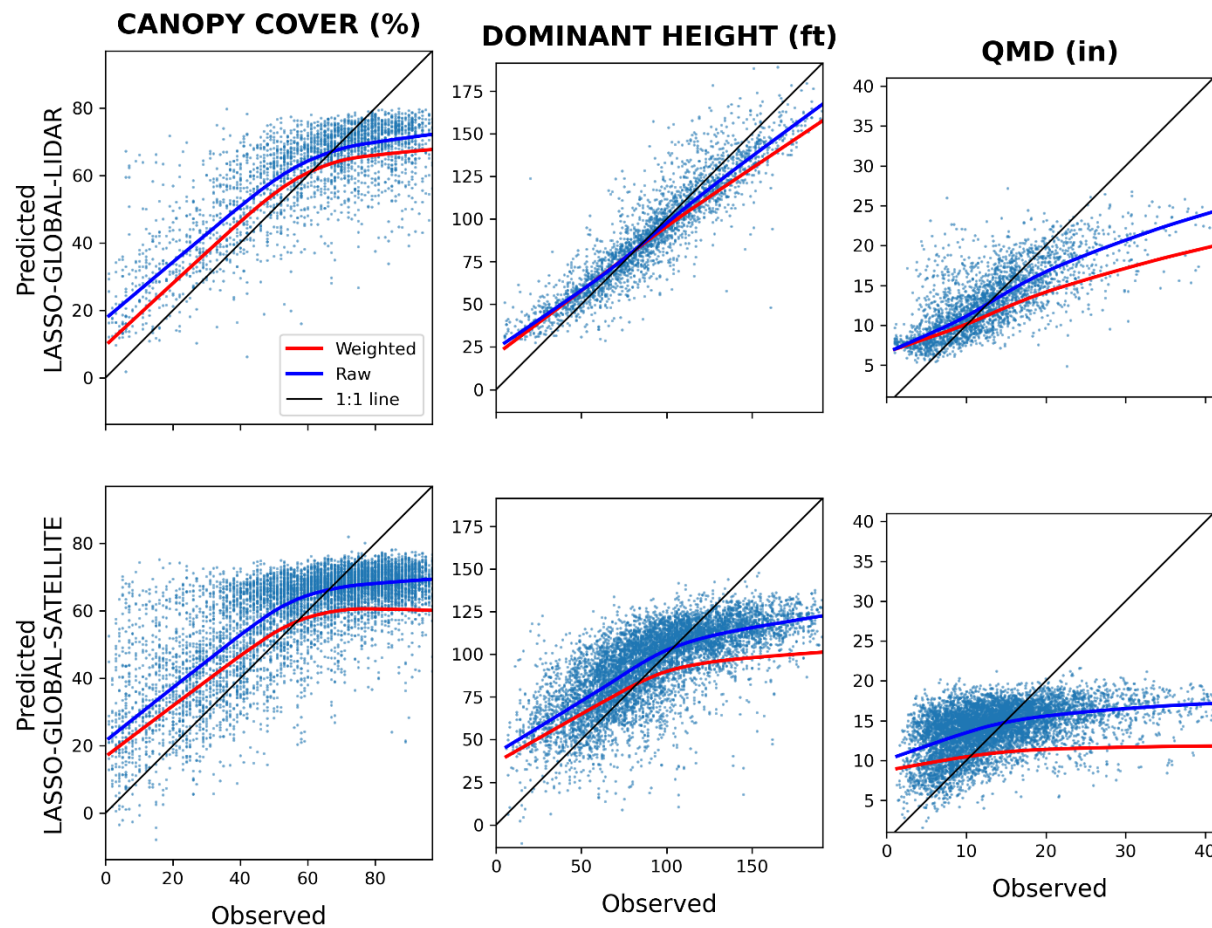


Figure 2.20. Correspondence between predictions and observations for linear satellite and lidar models (1 of 2).

These graphs show the predicted and observed values for Lasso models fit to lidar (top row) and satellite (bottom row) datasets. Values above illustrate predictions on data withheld from model training using five 80-20% folds. The blue line shows a locally weighted scatterplot smoothing function for the model trained on raw (unweighted) data, and red line for the model trained on weighted training data.

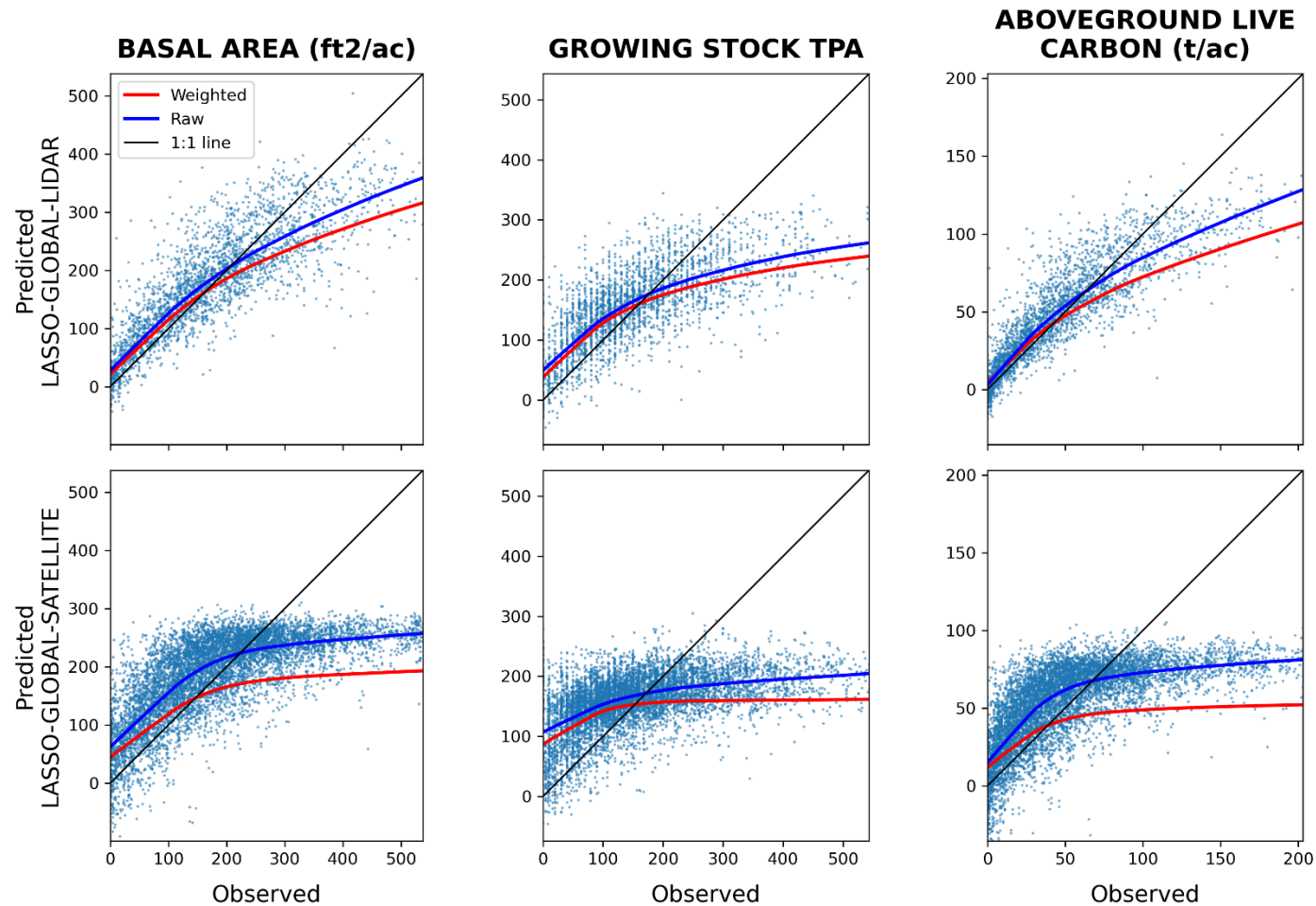


Figure 2.21. Correspondence between predictions and observations for linear satellite and lidar models (2 of 2). These graphs show the predicted and observed values for Lasso models fit to lidar (top row) and satellite (bottom row) datasets. Values above illustrate predictions on data withheld from model training using five 80-20% folds. The blue line shows a locally weighted scatterplot smoothing function for the model trained on raw (unweighted) data, and red line for the model trained on weighted training data.

Generalization capabilities, Insider vs. Global models

Applying the comparison threshold used by Fekety et al. (2018), I considered Global models practically equivalent to Insider models if the precision achieved by Global models was less than or equal to the average plus one standard deviation of the NRMSE of the Insider model estimated from five-fold cross validation. The outcomes of these comparisons among the nine ecoregions are shown in Table 2.5.

Table 2.5. Practical equivalence between Global and Insider models, NRMSE

Data Source	Model	Target					
		Cover	Dom Ht	BA	QMD	GS TPA	AGLTC
Lidar	ElasticNet	8/9	9/9	8/9	8/9	9/9	8/9
	Lasso	8/9	9/9	7/9	8/9	9/9	8/9
	kNN	8/9	9/9	8/9	9/9	9/9	8/9
	RF	8/9	9/9	8/9	9/9	9/9	7/9
	GBM	9/9	9/9	8/9	9/9	9/9	8/9
Satellite	ElasticNet	9/9	8/9	7/9	9/9	9/9	6/9
	Lasso	9/9	8/9	7/9	9/9	9/9	7/9
	kNN	9/9	9/9	7/9	6/9	9/9	7/9
	RF	9/9	9/9	8/9	6/9	9/9	8/9
	GBM	9/9	9/9	9/9	8/9	8/9	9/9

Notes: The values in the table show the number of ecoregions (out of 9 total) where a Global model showed performance deemed practically equivalent to the Insider model for that region. Global models showing NRMSE within the mean plus one standard deviation of the NRMSE for the Insider model are deemed practically equivalent. Target variables are canopy cover (Cover), dominant height (Dom Ht), live tree basal area (BA), quadratic mean diameter (QMD), growing stock trees per acre (GS TPA), and aboveground live tree carbon (AGLTC).

For the top-performing ML model (GBM), the few ecoregions where Global models showed average NRMSE exceeding Insider models were limited to estimates of aboveground carbon and basal area for lidar-based GBM models applied in the Willamette Valley, and for estimates of growing stock stem density from satellite-based GBM models applied in the Klamath Mountains. Visual inspection of the NRMSE model performance for GBM-Global and GBM-Insider models suggested this may have been an artifact of a single fold within five-fold cross-validation showing more extreme scores, whereas the scores from the other four folds for these

regions and target variables suggest practically equivalent or improved performance by the Global model compared with the Insider model (see Figure 2.10 and Figure 2.11).

When considering changes in bias when training data from outside a target region is added to a model, I observed that absolute bias did increase in several regions for several models, leading Global models to frequently show practically equivalent performance to Insider models (see Table 2.6).

Table 2.6. Practical equivalence between Global and Insider models, % Bias

Data Source	Model	Target					
		Cover	Dom Ht	BA	QMD	GS TPA	AGLTC
Lidar	ElasticNet	5/9	6/9	8/9	5/9	9/9	7/9
	Lasso	6/9	6/9	8/9	4/9	9/9	7/9
	kNN	7/9	9/9	8/9	7/9	9/9	7/9
	RF	8/9	9/9	7/9	8/9	9/9	6/9
	GBM	9/9	9/9	8/9	9/9	9/9	8/9
Satellite	ElasticNet	8/9	8/9	8/9	5/9	8/9	7/9
	Lasso	8/9	8/9	8/9	5/9	9/9	7/9
	kNN	9/9	8/9	7/9	7/9	9/9	6/9
	RF	8/9	8/9	6/9	6/9	8/9	6/9
	GBM	9/9	8/9	8/9	8/9	9/9	7/9

Notes: The values in the table show the number of ecoregions (out of 9 total) where a Global model showed performance deemed practically equivalent to the Insider model for that region. Global models showing % absolute bias within the mean plus one standard deviation of the % absolute bias for the Insider model are deemed practically equivalent. Target variables are canopy cover (Cover), dominant height (Dom Ht), live tree basal area (BA), quadratic mean diameter (QMD), growing stock trees per acre (GS TPA), and aboveground live tree carbon (AGLTC).

In the case of the best-performing ML model (GBM), bias did not change substantially between Insider and Global models for most ecoregions and targets. For lidar-based GBM models, only aboveground live carbon and live basal area in the Willamette Valley showed average % bias that exceeded the mean plus one standard deviation observed for the Insider models. Upon visual inspection, this difference appeared to be an artifact of single folds within the five-fold cross validation showing extreme values (see Figure 2.16). In the case of satellite-based GBM models, the following ecoregions and targets showed GBM-Global models with substantially different bias to GBM-Insider models: aboveground live carbon in the Blue Mountains and in the Northern Rockies; basal area in the Blue Mountains; QMD in the Puget Lowlands; and dominant height in

the Willamette Valley. I was unable to find a discernible pattern from these examples. I generally found that GBM-Global models had comparable precision and bias in the overwhelming majority of ecoregions and target variables.

Predictive improvements with increasing size and diversity of training data

I found significant improvements in predictive performance as models gained increasing volume and diversity of training data for all algorithms used. In general, the leap in model performance was the largest moving from Visitor models to Outsider models (see Figure 2.8 and Figure 2.9). Access to training data from within the target ecoregion had a much smaller influence over predictive performance compared to access to training data from other ecoregions across the broader Pacific Northwest outside the target ecoregion. This pattern was true for both model precision (e.g., NRMSE, also see Figure S2.28 and Figure S2.29) and bias (see Figure 2.14 and Figure 2.15).

Model transferability between ecoregions

When focusing on models trained in single ecoregions and applied in others, I observed that transferability of models varied dramatically depending upon the variable being predicted. Canopy cover models appeared to be most highly transferable between diverse ecoregions, while models for aboveground live tree carbon showed substantially worse performance with transfer across regions, for example (see Figure 2.22 and Figure 2.23). I observed increasing distance between regions for any variable (as measured by the Kolmogorov-Smirnov distance) corresponded to substantial increase in predictive error for the Lasso-regularized (linear) Visitor models, but did not appear to have a similar pattern with the GBM-Visitor models. This pattern of degrading transferability performance with increasing KS distance between regions is particularly apparent for linear models where the training region has fewer than 500 observations to learn from.

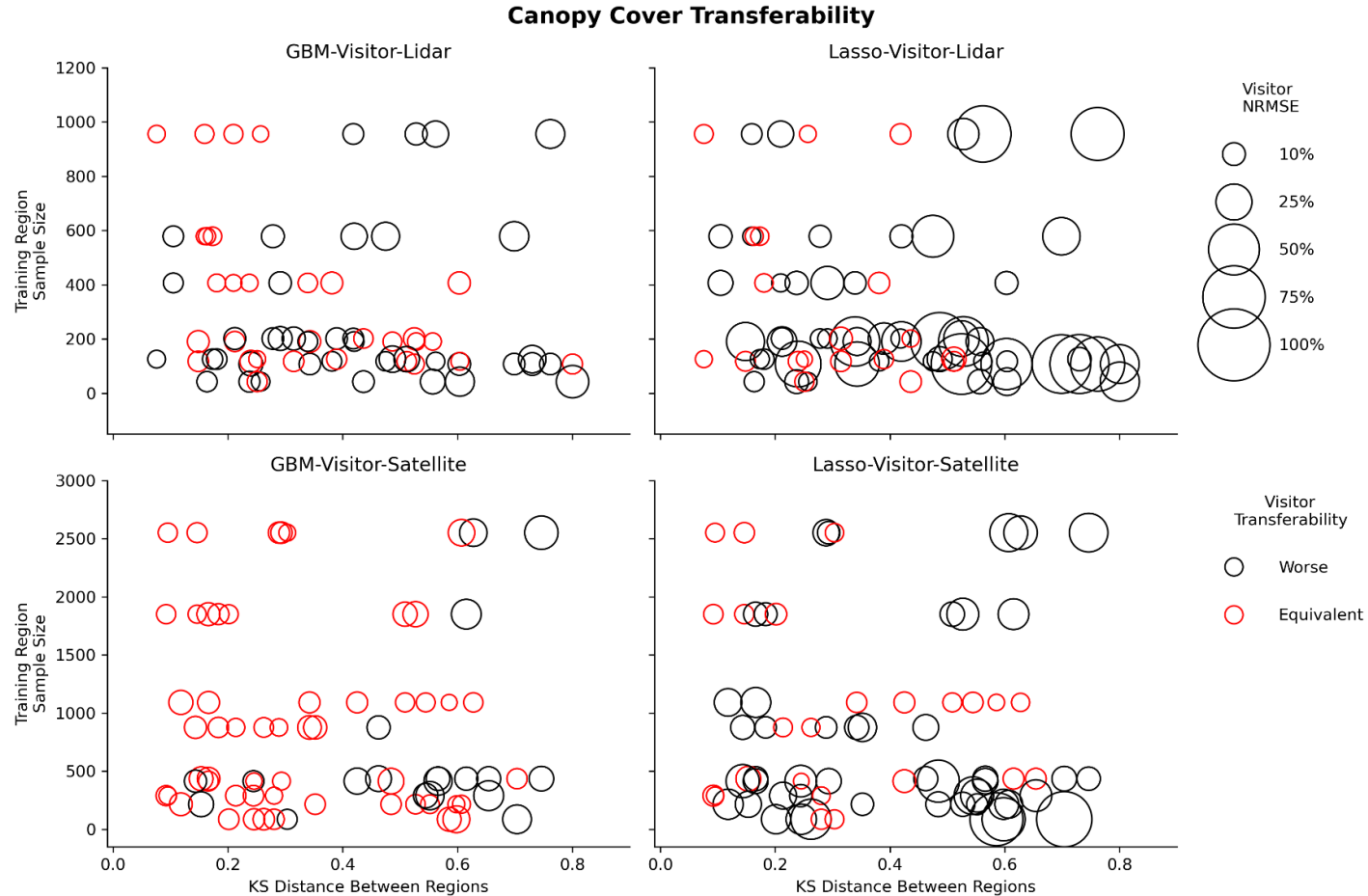


Figure 2.22. Visitor transferability for canopy cover predictions.

These graphs show the performance (in terms of NRMSE) of GBM and Lasso models trained in one region and applied in another based on the Kolmogorov-Smirnov distance between the training region and target region for canopy cover, and the number of samples in the training region. Visitor models that showed average NRMSE less than or equal to the mean + 1SD of the Insider model NRMSE for that target region are outlined in red.

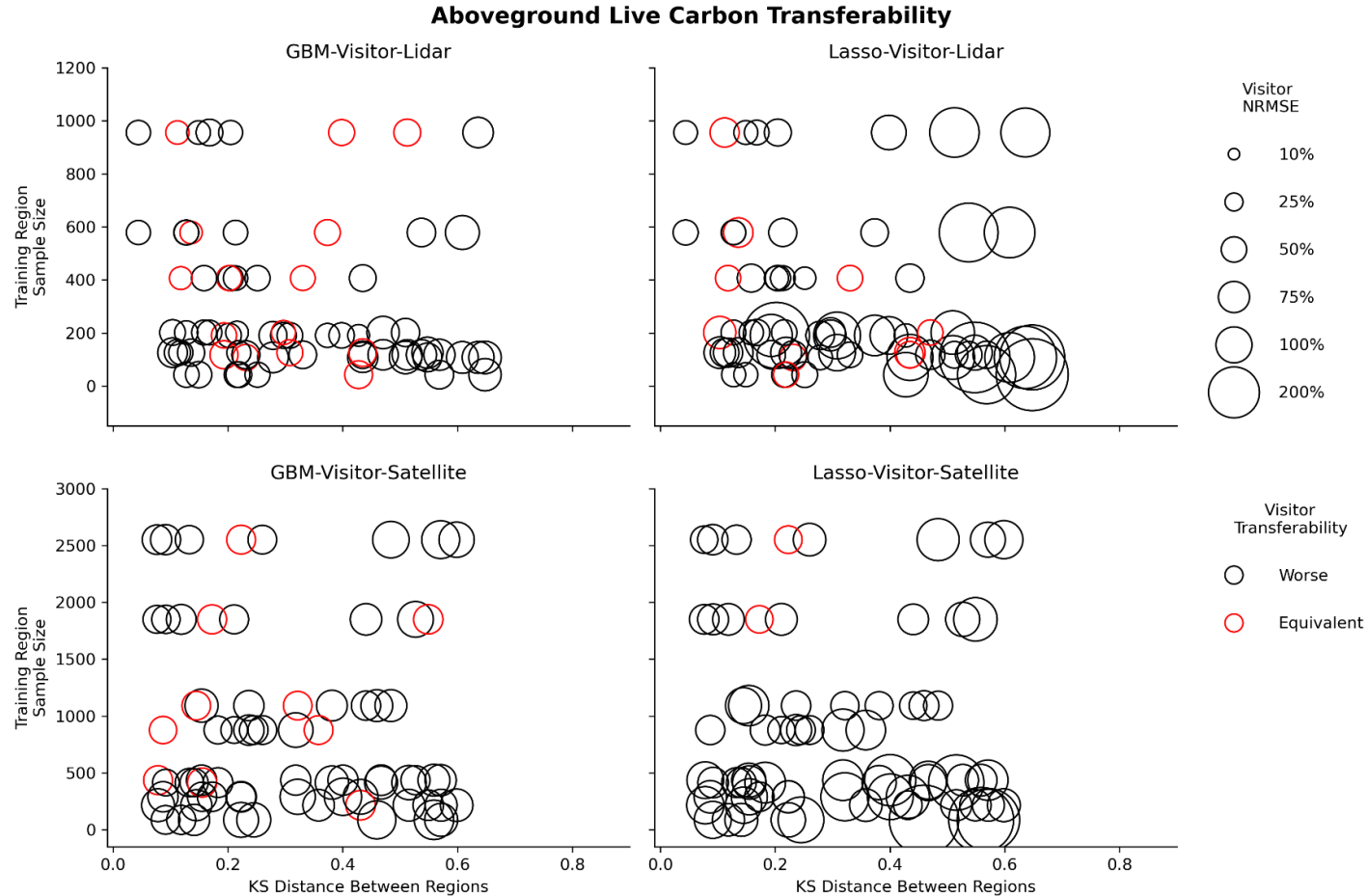


Figure 2.23. Visitor transferability for aboveground live carbon predictions.

These graphs show the performance (in terms of NRMSE) of GBM and Lasso models trained in one region and applied in another based on the Kolmogorov-Smirnov distance between the training region and target region for aboveground live carbon, and the number of samples in the training region. Visitor models that showed average NRMSE less than or equal to the mean + 1SD of the Insider model NRMSE for that target region are outlined in red.

Effects of sample weighting on model transferability

Focusing on the influence of sample weighting as a strategy to address domain shift between training datasets and target regions, I observed weak and inconsistent (in terms of direction) effects on model precision (e.g., NRMSE) of sample weighting when applied to GBM-Visitor models, and strong but also inconsistent (in terms of direction) effects on model precision when samples were weighted for Lasso-Visitor models. The effects of sample weighting on model precision did not appear to follow substantially different patterns between lidar-based and satellite-based models for a given algorithm (e.g., effects of weighting appeared similar for GBM-Visitor-Lidar as for GBM-Visitor-Satellite). Although the direction of effects of sample weighting were inconsistent on model precision, sample weighting generally appeared to have a negative effect on model performance in terms of bias for both Lasso-Visitor and GBM-Visitor models (see Figure 2.26 and Figure 2.27).

Across the range of each target variable, sample weighting had different effects. Sample weighting tended to reduce the overprediction bias commonly observed at the lower end of each structure variable's range, but to exacerbate the underprediction bias commonly observed at the upper end of the target variable's range. This pattern of weighting on bias across each variable's range appeared much more muted for GBM models than for Lasso models (see Figure 2.18 and Figure 2.19).

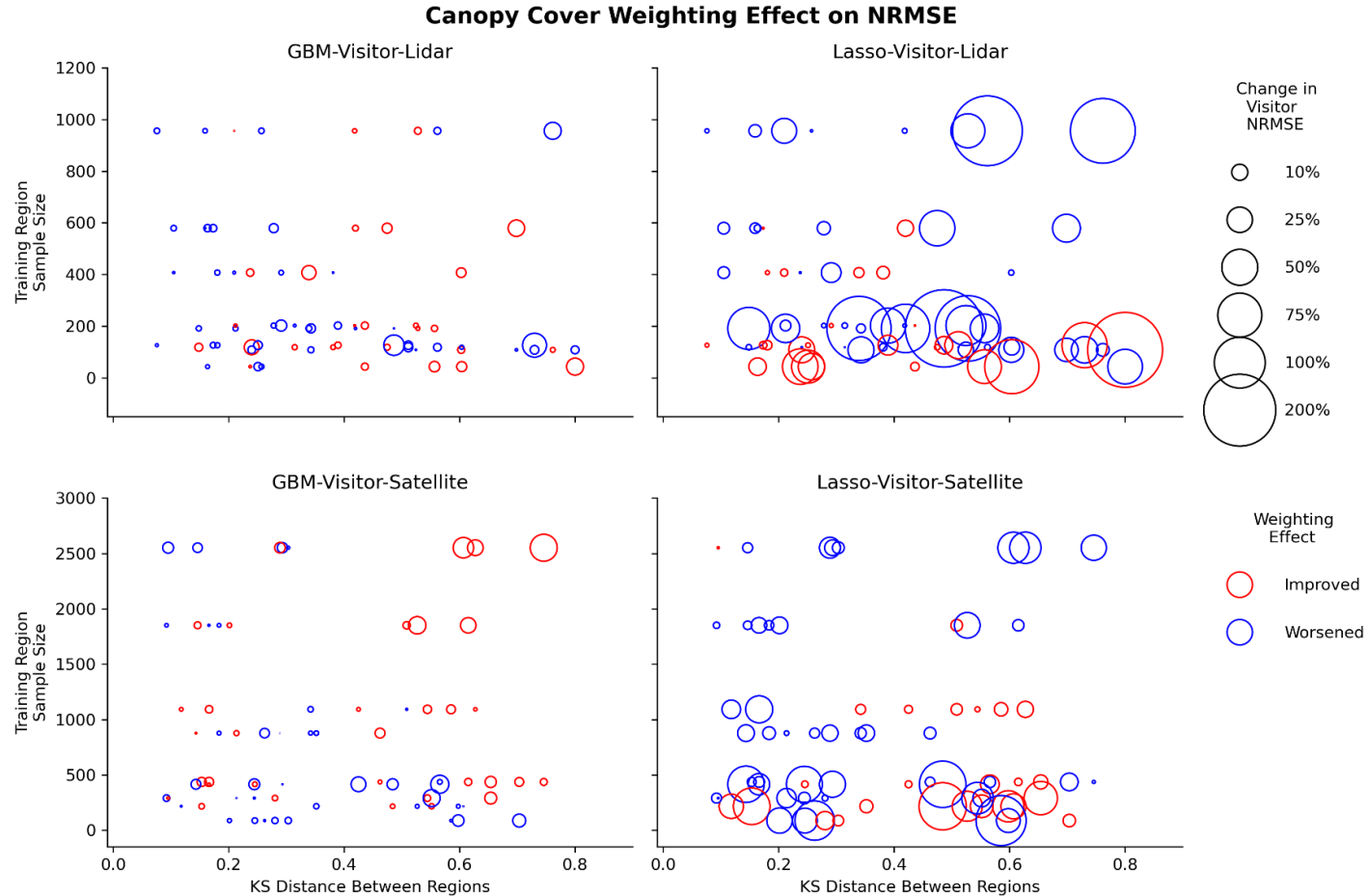


Figure 2.24. Effect of sample weighting on Visitor transferability for canopy cover predictions.

These graphs show the performance (in terms of NRMSE) of GBM and Lasso models trained in one region and applied in another based on the Kolmogorov-Smirnov distance between the training region and target region for canopy cover, and the number of samples in the training region. Visitor models that showed improved NRMSE are outlined in red, while those that worsened are outlined in blue. The size of the circle corresponds to the magnitude of change in NRMSE between weighted and unweighted models.

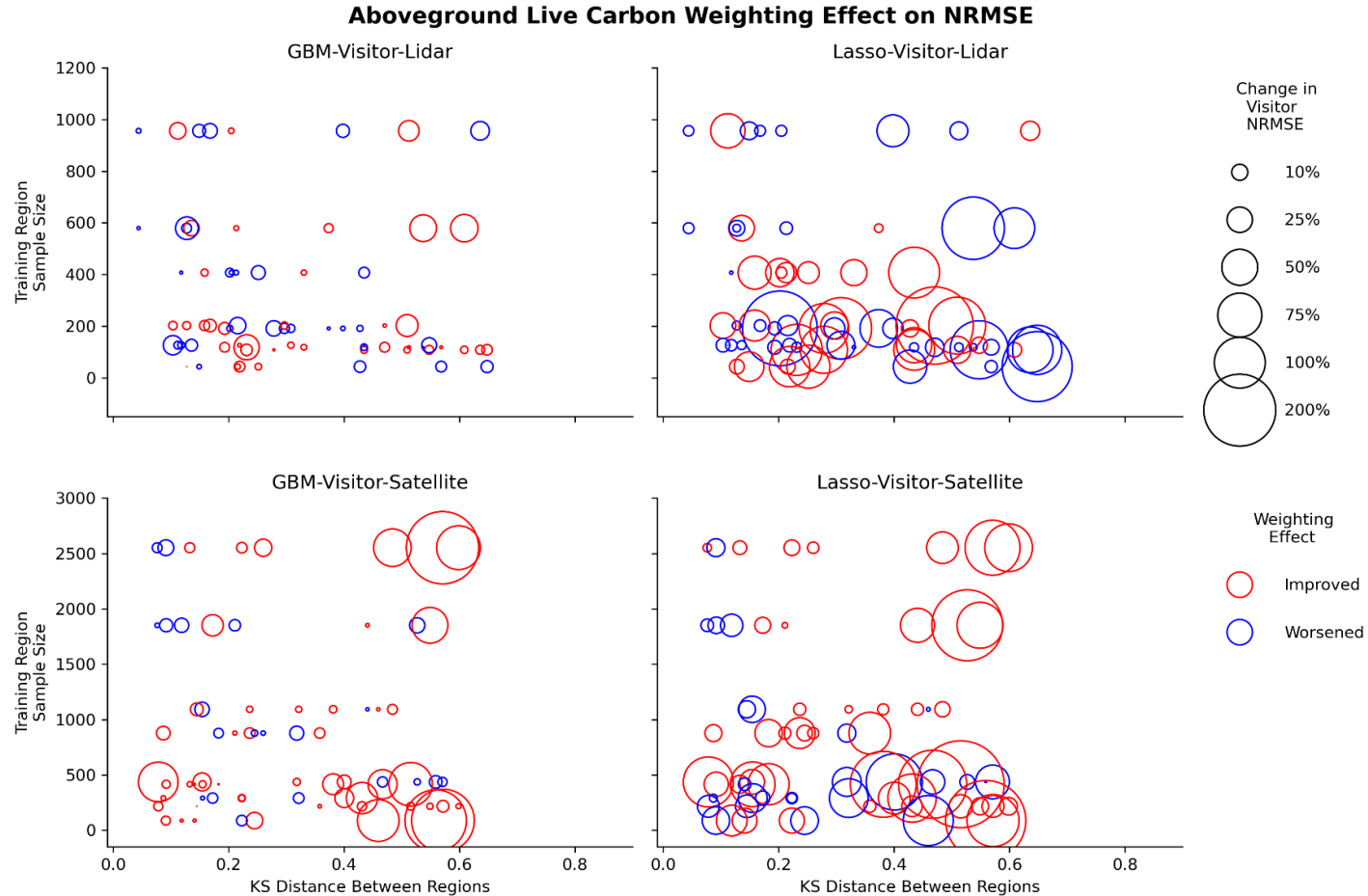


Figure 2.25. Effect of sample weighting on Visitor transferability for aboveground live carbon predictions.

These graphs show the performance (in terms of NRMSE) of GBM and Lasso models trained in one region and applied in another based on the Kolmogorov-Smirnov distance between the training region and target region for aboveground live carbon, and the number of samples in the training region. Visitor models that showed improved NRMSE are outlined in red, while those that worsened are outlined in blue. The size of the circle corresponds to the magnitude of change in NRMSE between weighted and unweighted models.

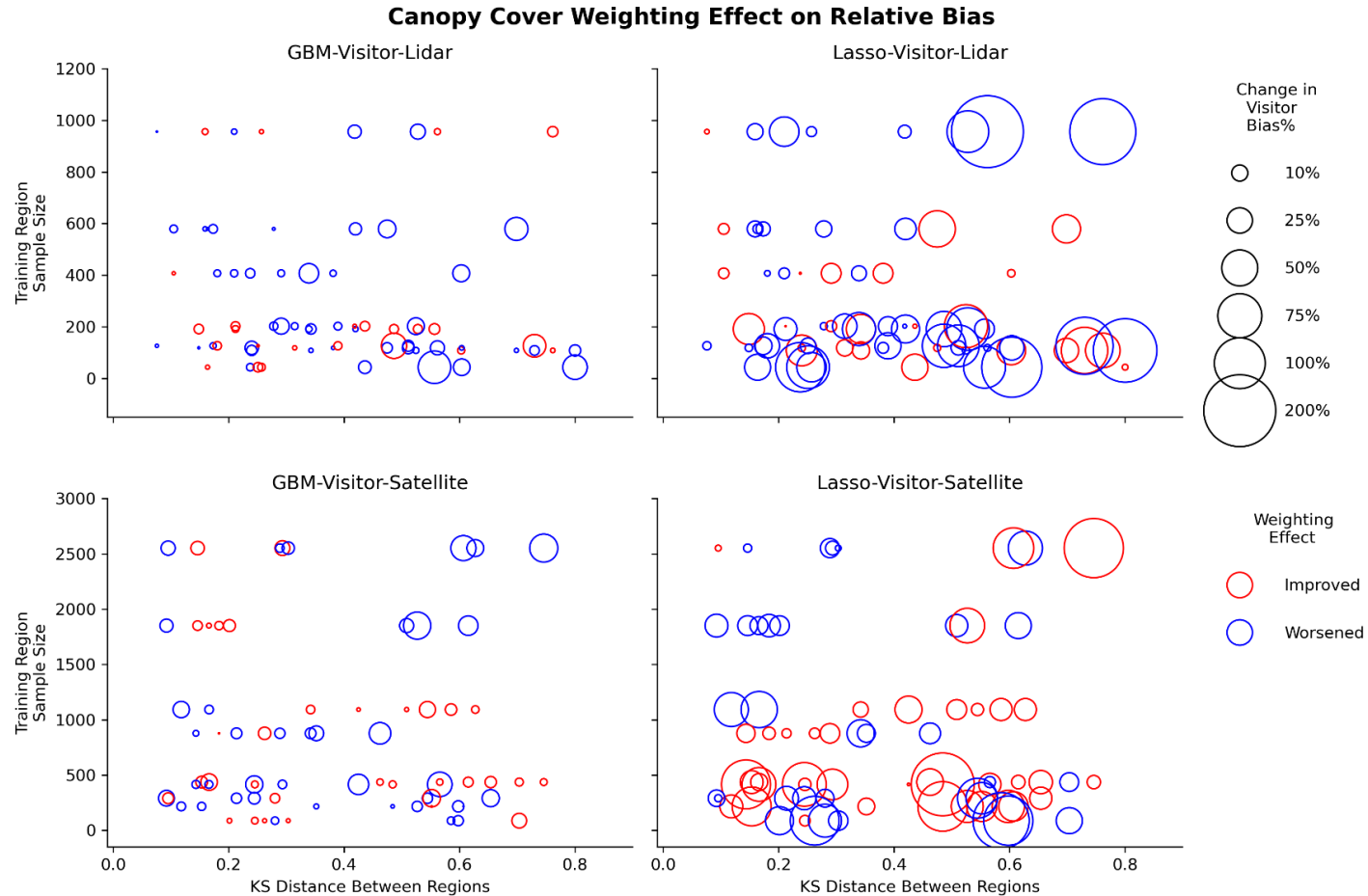


Figure 2.26. Effect of sample weighting on Visitor transferability for canopy cover predictions.

These graphs show the performance (in terms of relative bias) of GBM and Lasso models trained in one region and applied in another based on the Kolmogorov-Smirnov distance between the training region and target region for canopy cover, and the number of samples in the training region. Visitor models that showed improved % bias are outlined in red, while those that worsened are outlined in blue. The size of the circle corresponds to the magnitude of change in % bias between weighted and unweighted models.

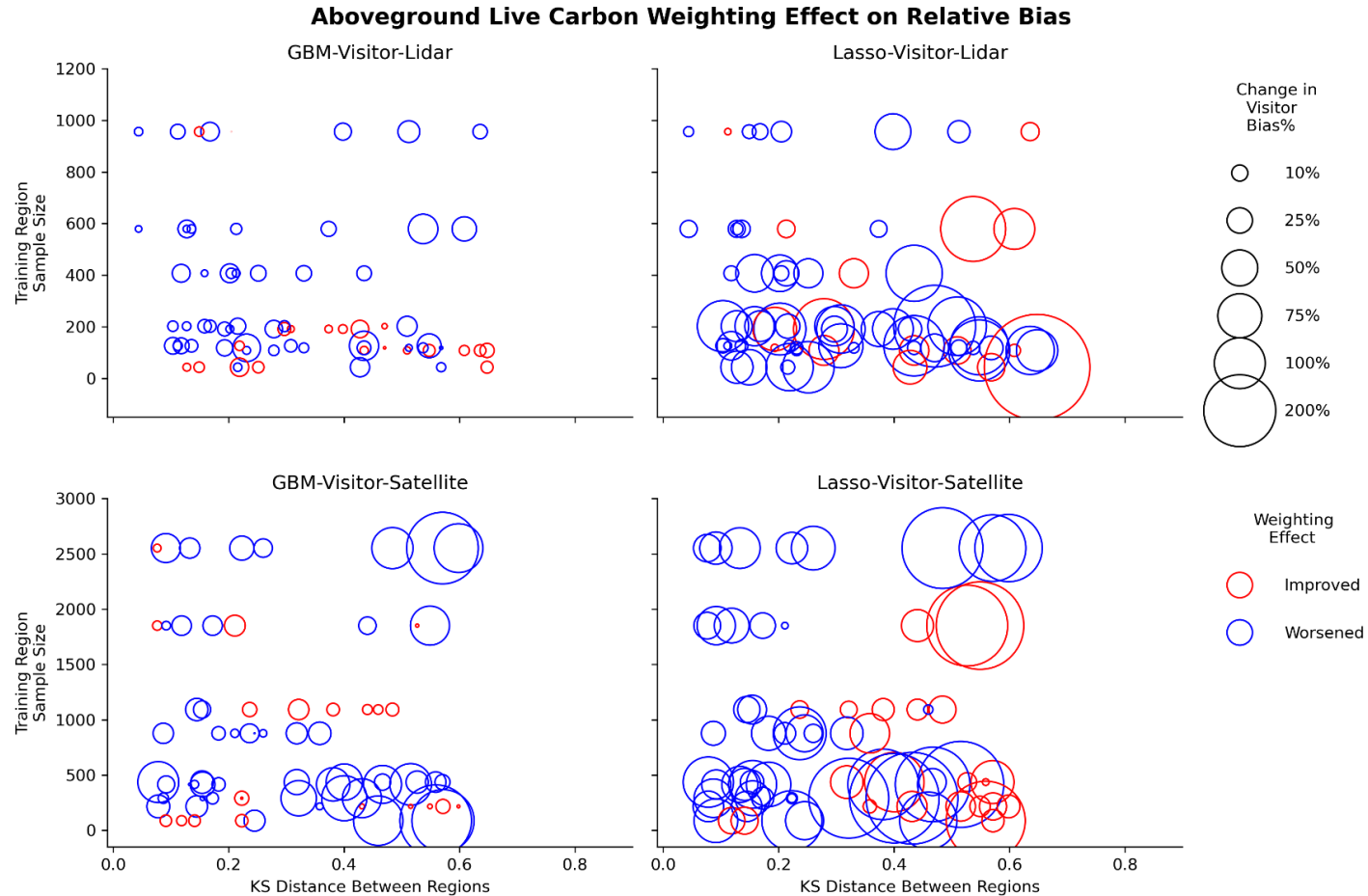


Figure 2.27. Effect of sample weighting on Visitor transferability for aboveground live carbon predictions.

These graphs show the performance (in terms of relative bias) of GBM and Lasso models trained in one region and applied in another based on the Kolmogorov-Smirnov distance between the training region and target region for aboveground live carbon, and the number of samples in the training region. Visitor models that showed improved % bias are outlined in red, while those that worsened are outlined in blue. The size of the circle corresponds to the magnitude of change in % bias between weighted and unweighted models.

2.4 DISCUSSION

The increased gain of predictive performance in terms of precision and bias seen with access to larger and more diverse datasets across the region as opposed to data within individual target ecoregions where a model is applied suggested that all algorithms investigated, including linear models, show potential benefits from being trained with data across multiple regions. This finding runs contrary to Ewijk et al. (2020) who observed increasing bias when data from the target region was added to a transferred model. I suspect this difference in finding may be related to the sample size and hold-out strategy used by Ewijk et al. (2020), who were conducting comparisons using only two regions with two research sites containing 397 and 92 plot observations in the two regions. In this study, the consideration of multiple ecoregions, access to a much larger sample size, and the addition of an Outsider scope in this study may allow us to speak more directly to the effects of access to multi-regional data distinct from the impact of accessing data from the target ecoregion.

RF and GBM models appear best-suited for these types of trans-regional applications (compared to linear models and kNN) due to the continued incremental gains in precision seen as these models are compared between Outsider > Insider > Global, with no substantial change in model bias. When considering both model precision and bias, I conclude that GBM models appear to offer the best overall performance compared to other algorithms, and that Global models offer performance that typically meets or exceeds the performance offered by models trained in single ecoregions.

Given the substantially larger sample sizes available from the compilation of plot-level observations across agencies and ecoregions and apparent transferability capacity of non-linear ML models, I suggest that ML models such as GBM fit to multi-regional datasets involving thousands of samples covering the diversity of forest structures observed across a bioregion are likely to mitigate the need for significant new field campaigns to collect plot observations for developing correlative models with remote sensing in the Pacific Northwest for most forest conditions. The compilation of multi-agency remote sensing datasets is likely to provide a suitable basis for most forest monitoring and mapping needs such that additional field measurement can and should be focused on increasing sample sizes for forest conditions that the models show

poorest performance, particularly in areas at the lower and higher ends of the ranges of forest structure conditions, including forest conditions with more sparse and open canopies.

Among the various performance metrics and consideration of both precision and bias, GBM appears to be a consistently top-performing model that offers desirable indications of transferability and generalization across and among regions. Similar to Cosenza et al (2020), but contrary to my initial expectations, I observed kNN models generally showed worse performance than either linear models or tree-based models (GBM and RF). As expected, the RF and GBM models showed improved performance relative to the linear models, particularly in terms of predictive precision. The observations of substantial negative biases for kNN and RF models raise some concern given the widespread use of similar models across the Pacific Northwest for regional analysis, although some bias-correction methods recently introduced (Hudak et al., 2020) using comparisons with FIA population estimates at county- and larger geographic extents may help mitigate this issue substantially.

Consistent with my hypothesis, I found that lidar models showed the highest predictive performance in terms of precision and bias. Across regions, I generally observed predictive performance of Global and Insider lidar-based models are likely to satisfy the levels of precision needed for intensive and precision forestry use cases based on feedback from beta-testing with family forest owners and consulting foresters in Oregon, and from an advisory team including family forest owners, land trust managers, consulting foresters, and state and federal managers of forest conservation and planning programs for non-industrial private forest owners.

In the ranges of forest height and average tree diameters where changes in boardfoot volume and value used to drive commercial timber planning decisions are particularly sensitive to upward or downward biases, the predictive performance of the best-performing satellite models appears unlikely to meet the needs for intensive forestry planning or forest planning that is particularly concerned with boardfoot volume and value. In contrast, however, the satellite-based estimates of basal area, aboveground live carbon, and canopy cover generally seem to be sufficiently precise and unbiased in the range of forest conditions commonly observed among non-industrial forest owner properties and may be suitable for forest conservation and stewardship planning in these extensive management contexts. As these maps and predicted forest attributes are integrated into a publicly-accessible web application (<http://landmapper.ecotrust.org>), we

expect to continue gathering user feedback to identify forest conditions types that are not being adequately modeled, and to update models using an iterative approach.

The systematic biases observed with satellite data raise a substantial concern regarding the ability for satellite-based models to consistently detect and monitor late successional forest conditions in the upper ranges of basal area and carbon stocking. For public forest agencies where substantial regulatory and management attention are directed these forest types, lidar-based models may be necessary to confidently assess and monitor them. Initial feedback from beta-testers and advisors suggest estimates of average diameter and growing stock stem density from satellite-based models appear borderline for extensive or non-industrial use cases at property- and stand-scales, but may be suitable for larger landscape planning efforts where decisions are not particularly sensitive to these variables.

This study utilized canopy cover derived from allometric equations (within the Forest Vegetation Simulator) applied to inventory data as an estimate of canopy cover in the field. Considering canopy cover is one of the forest attributes most directly related to what remote sensing instruments “see”, the comparable amount of error seen in models trained from lidar and from satellite data sources suggests that underlying error or bias in the allometric equations may be substantially confounding estimates of the performance of the predictive models. Since the active sensing of lidar is known to detect the occurrence of vegetation very precisely, it would seem preferable to adopt models that incorporate lidar-derived estimates of canopy cover as more reliable than allometric estimates from inventory data. In particular, it has been documented that FVS equations for estimating canopy cover, particularly regarding adjustments to account for tree spacing and crown overlap, tend to underestimate canopy cover in dense closed canopy conditions (Gray et al., 2021). I suspect this systematic underestimation of dense canopy cover by allometric models may be a major contributor to the saturating effect observed in predictions of canopy cover for both lidar- and satellite-based models.

Sample weighting did not appear to substantially alter the performance of GBM models, but did have significant effects on the performance of linear models in terms of magnitude. Mauro et al. (2021) found that pre-calibration substantially affected linear model bias but was minimal for semi-parametric models (RF models with random effects added for each field plot and lidar acquisition). However, in contrast to Mauro et al. (2021), I observed sample weighting to be much more likely to substantially increase the bias of linear models transferred to other regions. I

interpreted this muted effect of sample weighting to be related to the ability for RF and GBM models to build detailed non-linear representations of relationships between predictor and target variables, limiting the influence of sample weighting or pre-calibration attempts as the training dataset grows beyond a few hundred observations. That being said, I did not characterize regional-scale effects of model weighting, which may be more subtle and relevant to consider for use cases at regional extents. It is not clear, for example, whether sample weighting may mitigate the need for regional bias correction following models similar to Hudak et al. (2020).

2.5 FUTURE STUDY

The extension of this study to consider the effects of sample weighting in terms of county- and larger-scale population estimates would be relatively modest, but important step toward justifying the use of predictive maps like these for landscape-level planning and monitoring. Similarly, it seems likely that certain forest structure attributes will be more strongly influenced by plots of varying sizes, such that a sensitivity analysis into the amount of weight that should be given to smaller plots be reduced compared to larger plots. Intuitively, both stem counts and average diameter seem like structure attributes that may be more sensitive to the choice of plot size than the other structural variables considered in this study, and a preference for larger plots or incorporation of uncertainty in training data may be worthwhile for improving predictive performance among small diameter stands and stands with relatively low stem densities.

The combined recognition that growing the number and diversity of training plots appears to significantly enhance the performance and generalization ability of predictive models seems well-suited to the application of a semi-supervised approach where lidar-based models could be applied across entire lidar acquisitions, with satellite-based models subsequently trained based on the lidar-based predictions. Luther et al. (2019) demonstrated a satellite model trained in this “indirect” way showed improved performance on a validation dataset compared to a model trained directly on a smaller dataset containing only the actual field observations. This approach would also leverage the ability for lidar to provide a more confident estimate of canopy cover than the application of allometric equations to inventory data, and provide opportunities for satellite-based models to be trained using a more robust representation of canopy cover samples at both low- and high-ends of canopy cover that are less commonly collected in inventory plots or reported from allometric equations adjusted for crown overlap. In addition to increasing the number of training

samples for satellite-based models by several orders of magnitude compared to direct training on field observations, this semi-supervised approach also opens up the opportunity to explore sample weighting more deeply to try and further mitigate prediction biases across the range of forest conditions.

Finally, most forest structure models developed and presented in the literature to date rely on models that fail to integrate spatial patterns in forest structure directly into predictions. All the studies referred to here, and this study as well, have not employed the latest generation of ML models which have shown some of the most dramatic improvements in both classification and regression tasks compared to earlier ML algorithms: Convolutional Neural Networks (CNNs). Recognizing the task of predicting forest structure is fundamentally one that involves patterns emerging at multiple scales, CNNs and computer vision approaches more generally seem to be better tailored for this type of modeling problem. CNNs would also appear well-suited for recognizing issues of varying plot sizes as well, and to models that would allow for methods of “data augmentation” to account for uncertainty in plot location and co-registration of field attributes with raster pixels of varying resolutions.

One of the major bottlenecks with fitting deep learning models including CNNs is the need for large volumes of training data and the need for the processing of data to maintain spatial relationships. In this study, for example, I summarized all lidar- and satellite-based features into a single summary statistic for each plot. To employ CNNs, the training data need to be formatted to allow models to consume images or arrays that explicitly retain spatial variation in the predictor and target variables. The semi-supervised approach mentioned above to employ predictive models across lidar acquisitions and then to train computer vision models using satellite imagery may be well-suited for generating large sample sizes with relatively limited bias in spatial texture. As an increasing amount of forest inventory data collected by public agencies with imprecisely-located plot locations is likely to emerge in the future, I also expect the application of small area estimation techniques such as Fay-Herriot models to be of increasing relevance (Temesgen et al., 2021). Along with the ability to simultaneously control for biases that may emerge at larger spatial scales, the potential for area-based models to incorporate variable radius plot data and other common forms of stand-based inventory data that are likely to outnumber confidently-located plots by several orders of magnitude represents a significant opportunity for improving multi-scale

representations in models used to predict forest structure and composition, and our ability to monitor and learn from forest conditions and dynamics over time.

2.6 CONCLUSION

This study effectively demonstrates the potential for the compilation of diverse public datasets to meet forest mapping needs for a variety of use cases. In particular, this approach clearly offers a unique opportunity to deliver actionable information about forest conditions across broad geographic regions for landowners who have generally benefited the least from recent advancements in remote sensing, computation, and predictive modeling. This data compilation and predictive modeling holds the potential to significantly reduce the cost, time, and effort required to conduct stand examinations and collect inventory data that is often a substantial bottleneck in the development of forest management plans.

I bolster earlier studies that demonstrate the potential for ML models to generalize effectively and transfer across multiple regions. As the sample size and diversity of forest conditions represented in publicly available datasets continues to grow, I expect concerns regarding domain shift and transferability to be concerns in fewer and fewer use cases. For the Pacific Northwest, I conclude that a sufficiently large and diverse set of forest conditions have now been sampled and can be compiled to allow for the development and application of models across Oregon and Washington by a growing range of forest managers and stakeholders who have previously been constrained by limited resources or access to collect or analyze remote sensing and field inventory data. Over time, I expect this transition to increasingly support a strategic shift towards targeted field measurements to correct models rather than to reinforce fundamental representations of forest structure and remote sensing attributes.

2.7 SUPPLEMENTARY MATERIALS

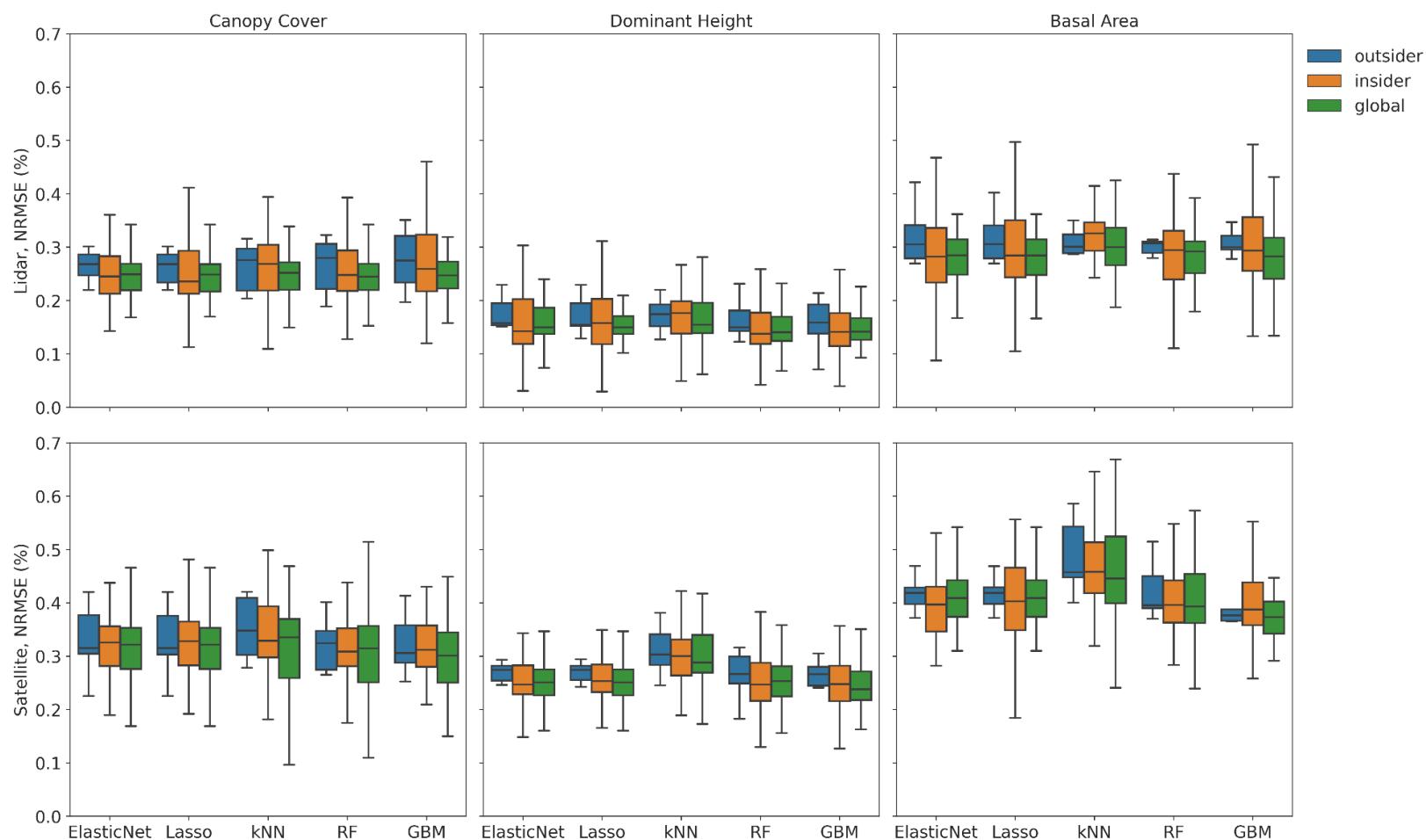


Figure S2.28. Improved predictive performance (Normalized Root Mean Square Error, NRMSE) with increasing training data for three structure attributes (1 of 2).

The middle line in each box displays the median, the upper and lower edges the 25th and 75th percentiles, and whiskers for 5th and 95th percentiles. Individual observations outside these ranges are shown as points. For Global and Insider models, each box summarizes five cross-validation scores across nine ecoregions (45 values); and for Outsider models, each box summarizes nine scores (one Outsider model applied to each of nine ecoregions).

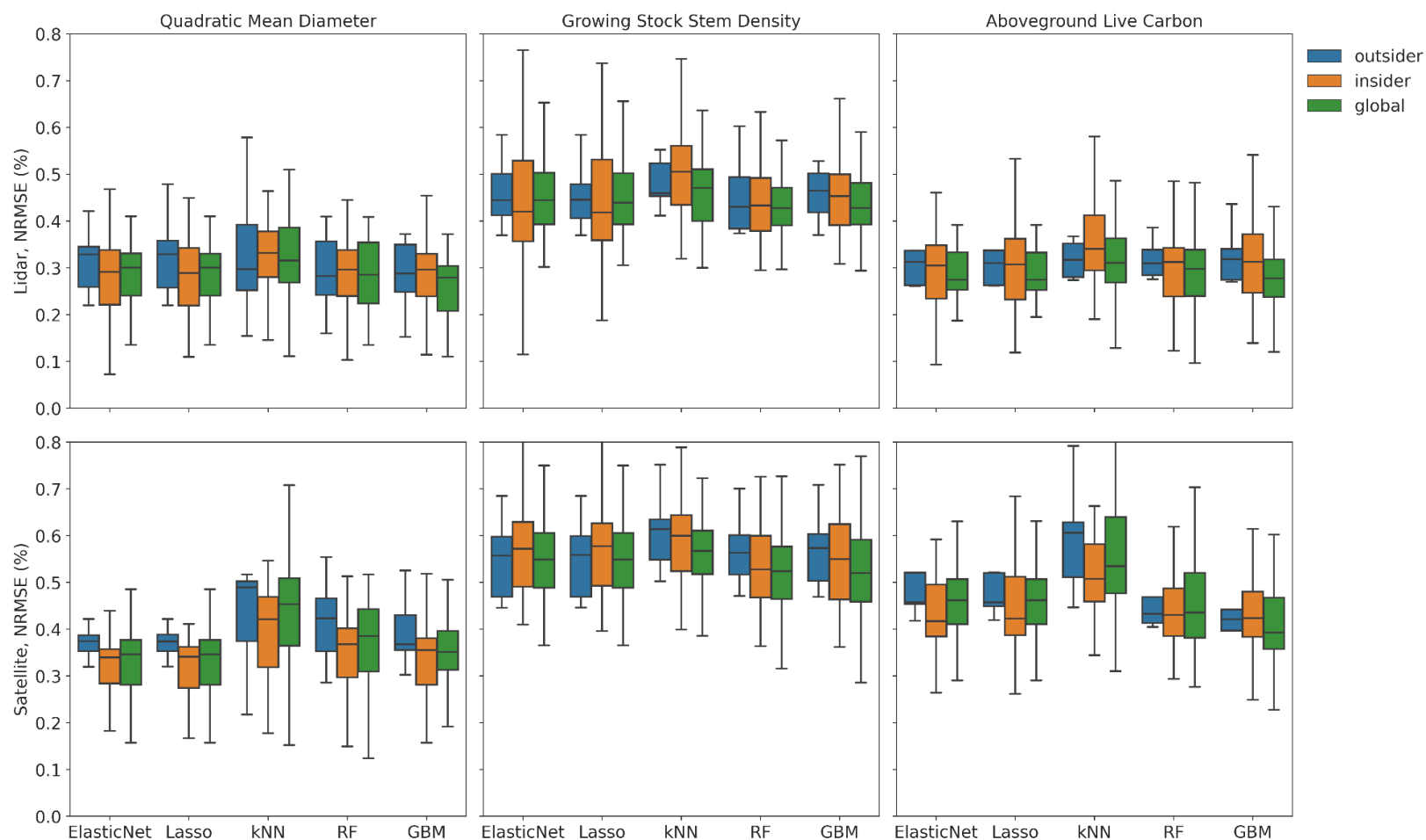


Figure S2.29. Improved predictive performance (Normalized Root Mean Square Error, NRMSE) with increasing training data for three structure attributes (2 of 2).

The middle line in each box displays the median, the upper and lower edges the 25th and 75th percentiles, and whiskers for 5th and 95th percentiles. Individual observations outside these ranges are shown as points. For Global and Insider models, each box summarizes five cross-validation scores across nine ecoregions (45 values); and for Outsider models, each box summarizes nine scores (one Outsider model applied to each of nine ecoregions).

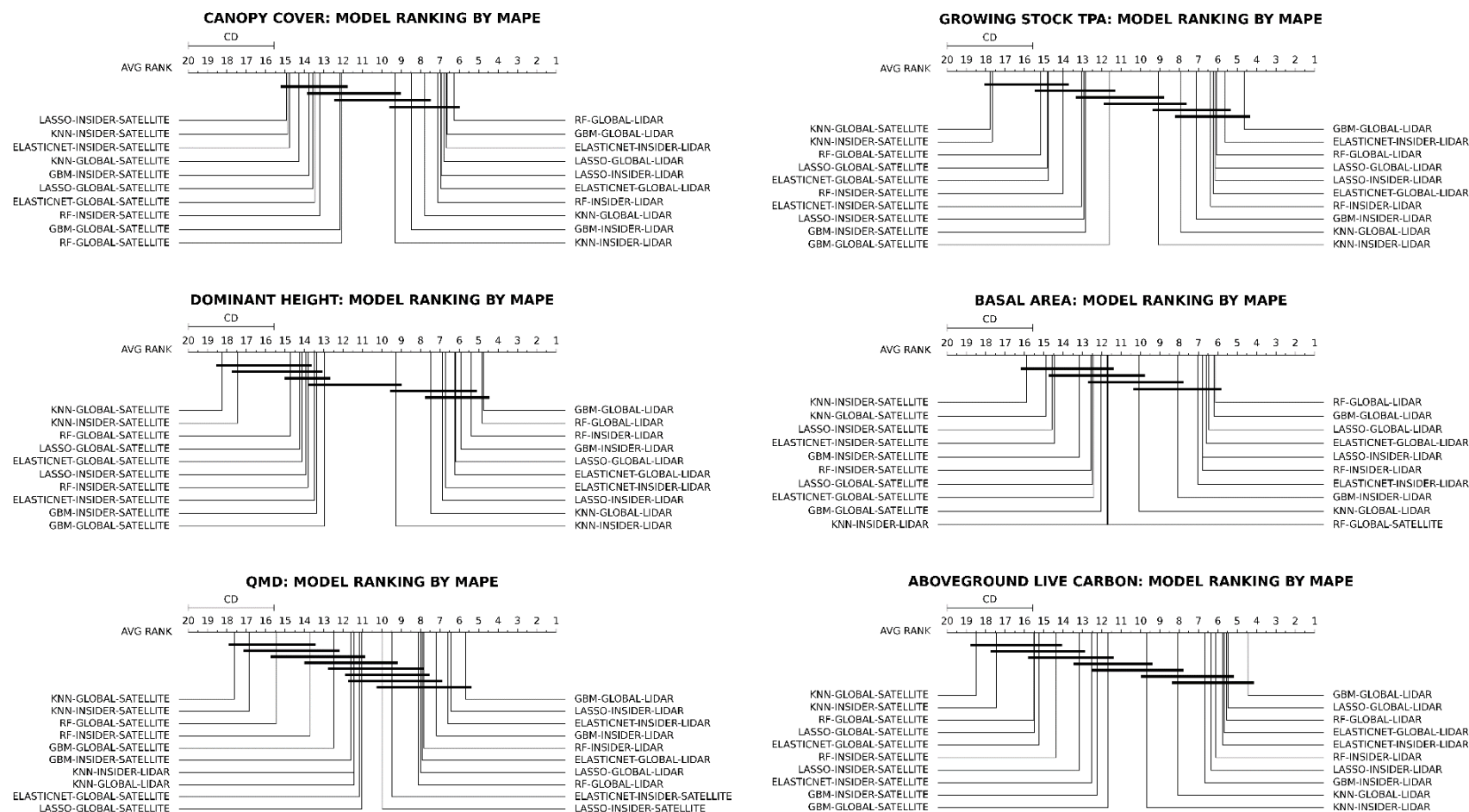


Figure S2.30. Rankings of forest structure predictive models based on mean average percent error.

Each graph displays the average ranking of models based on weighted training data for each combination of algorithm, scope (insider or global) and source (lidar or satellite) across all ecoregions following Demšar (2006). Models that are not statistically significantly different from each other are displayed as crossing the same horizontal black bar below the rank scale.

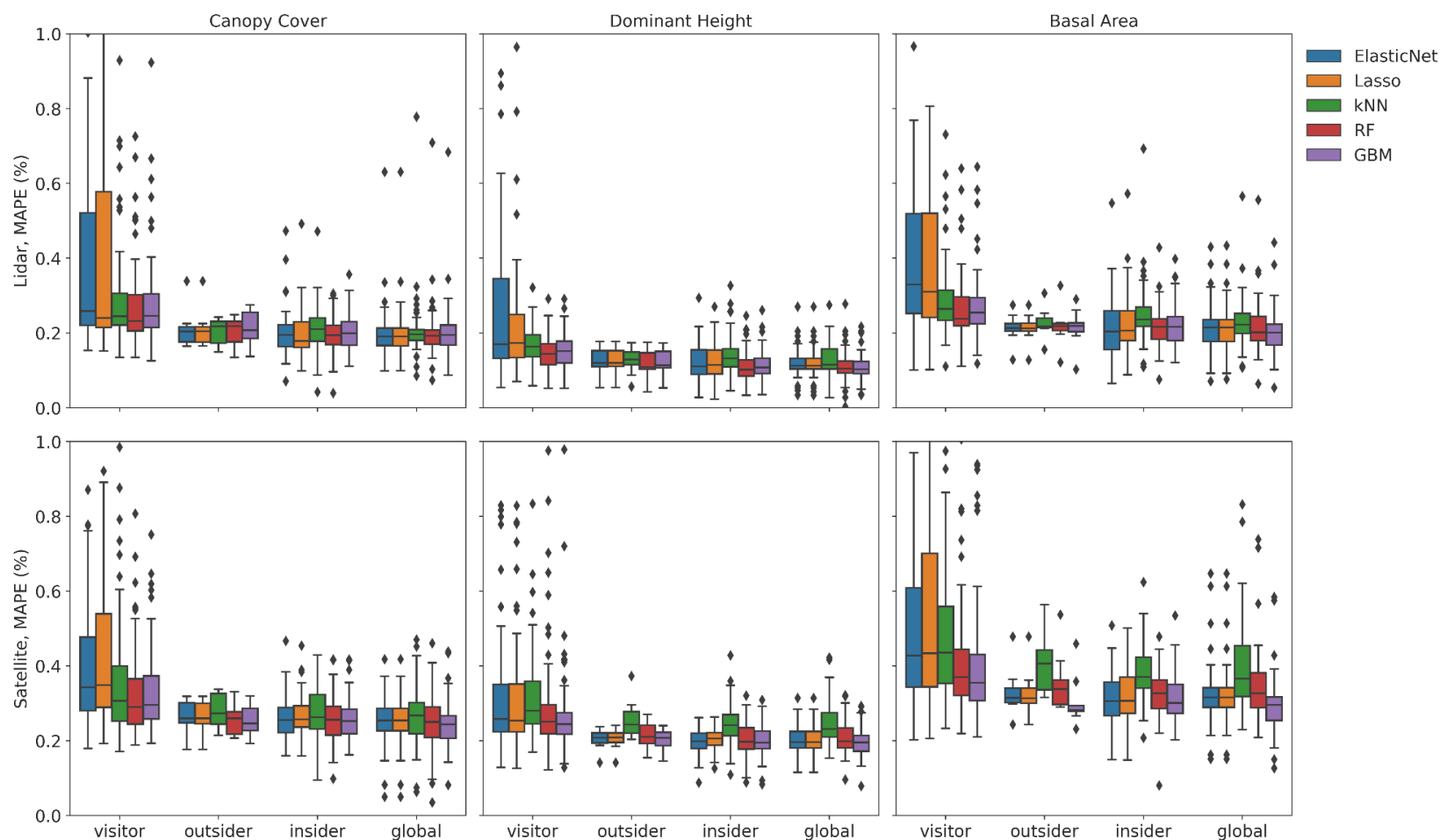


Figure S2.31. Predictive performance (Mean Absolute Percent Error, MAPE) of lidar- and satellite-based models on three forest structure attributes (1 of 2).

The middle line in each box displays the median, the upper and lower edges the 25th and 75th percentiles, and whiskers for 5th and 95th percentiles. Individual observations outside these ranges are shown as points. For Global and Insider models, each box summarizes five cross-validation scores across nine ecoregions (45 values); for Visitor models, each box summarizes 72 scores of algorithms trained in one of nine ecoregions applied in another (for each of nine regions, there are eight Visitor models scored once); and for Outsider models, each box summarizes nine scores (one Outsider model applied to each of nine ecoregions).

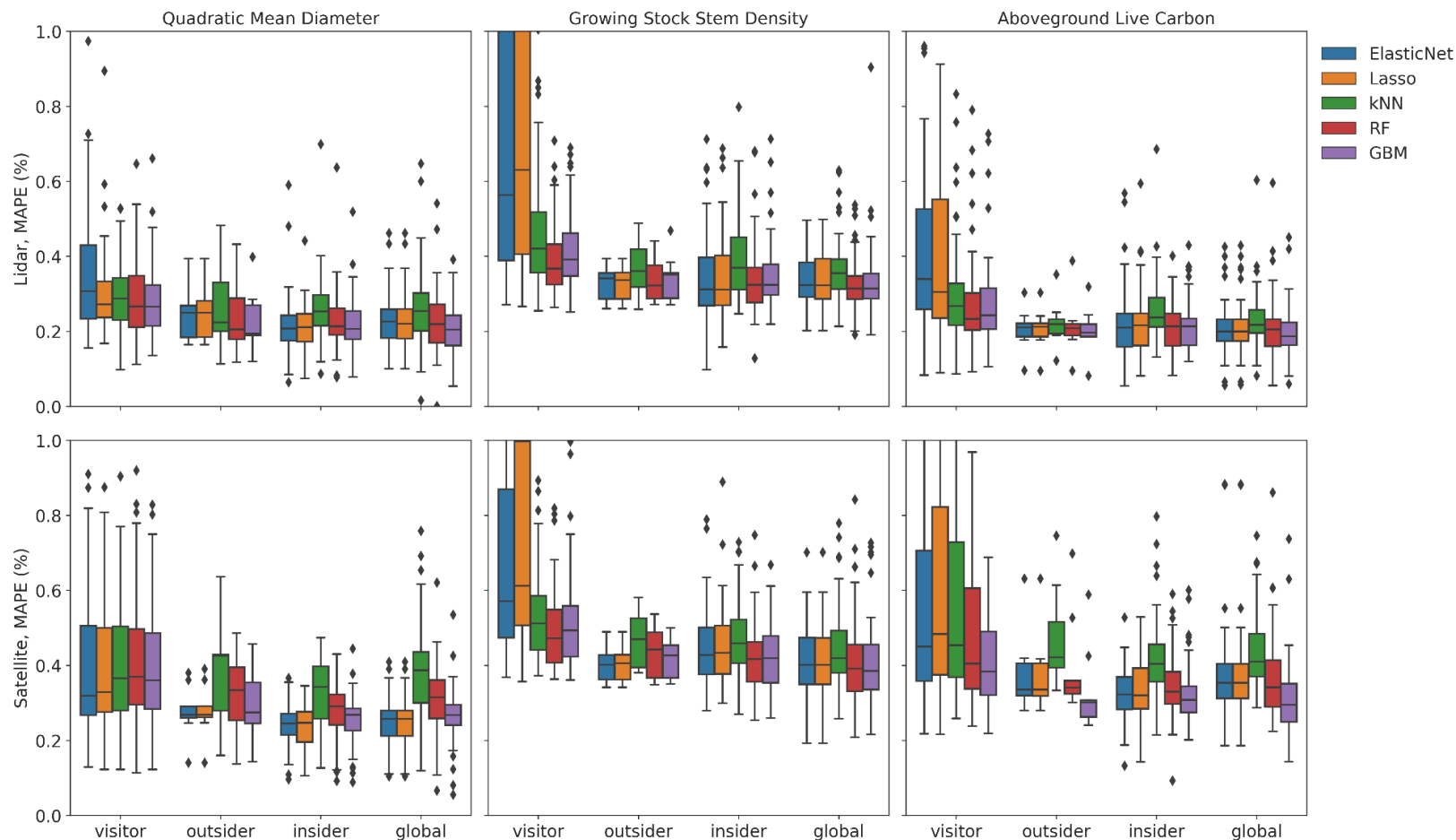


Figure S2.32. Predictive performance (Mean Absolute Percent Error, MAPE) of lidar- and satellite-based models on three forest structure attributes (2 of 2).

The middle line in each box displays the median, the upper and lower edges the 25th and 75th percentiles, and whiskers for 5th and 95th percentiles. Individual observations outside these ranges are shown as points. For Global and Insider models, each box summarizes five cross-validation scores across nine ecoregions (45 values); for Visitor models, each box summarizes 72 scores of algorithms trained in one of nine ecoregions applied in another (for each of nine regions, there are eight Visitor models scored once); and for Outsider models, each box summarizes nine scores (one Outsider model applied to each of nine ecoregions).

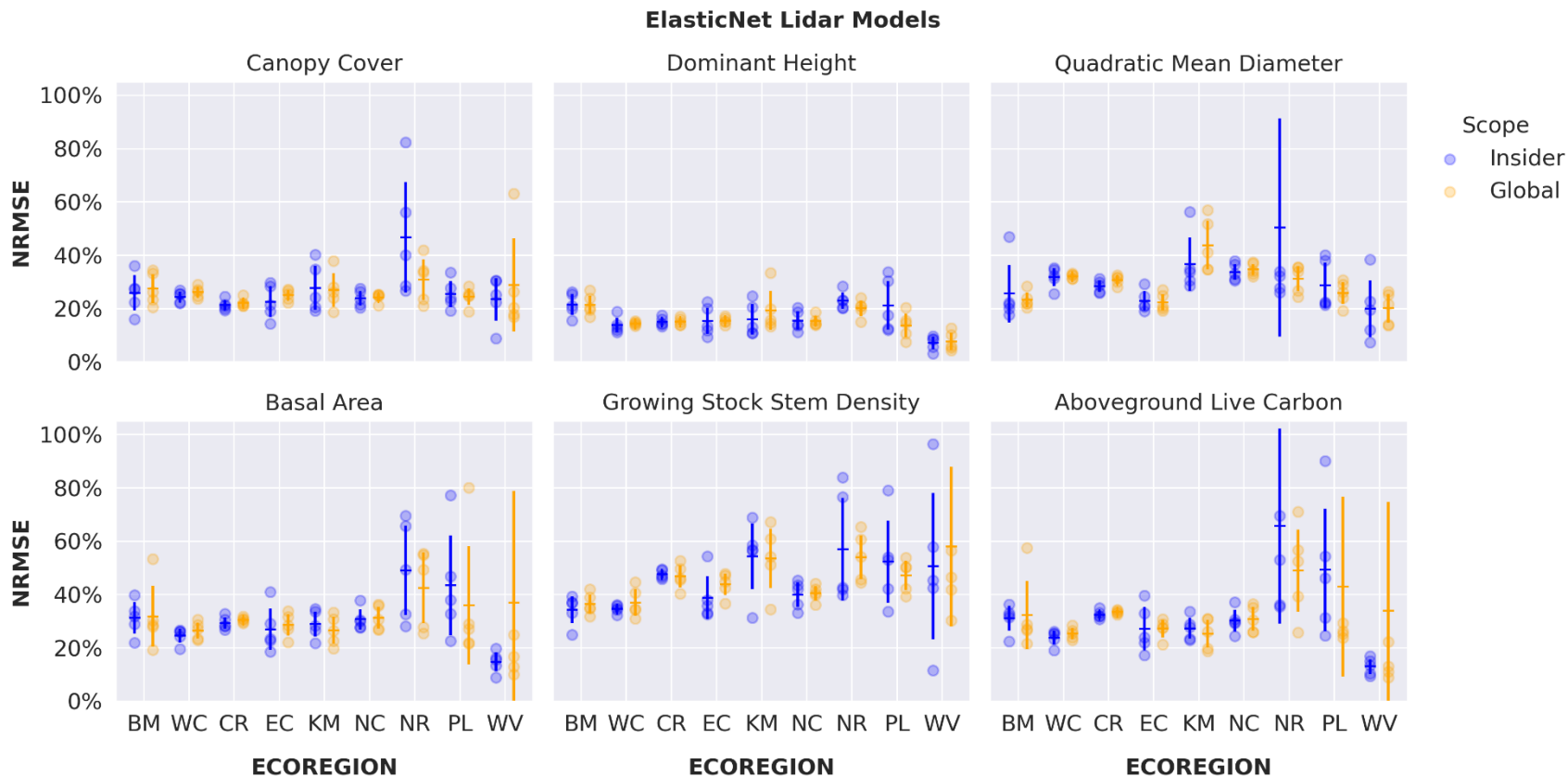


Figure S2.33. Performance (Normalized Root Mean Square Error, NRMSE) of ElasticNet-GLOBAL and ElasticNet-INSIDER lidar models across all ecoregions.

These graphs show the scores from five rounds of cross-validation for the ElasticNet Insider and Global models trained from lidar data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.

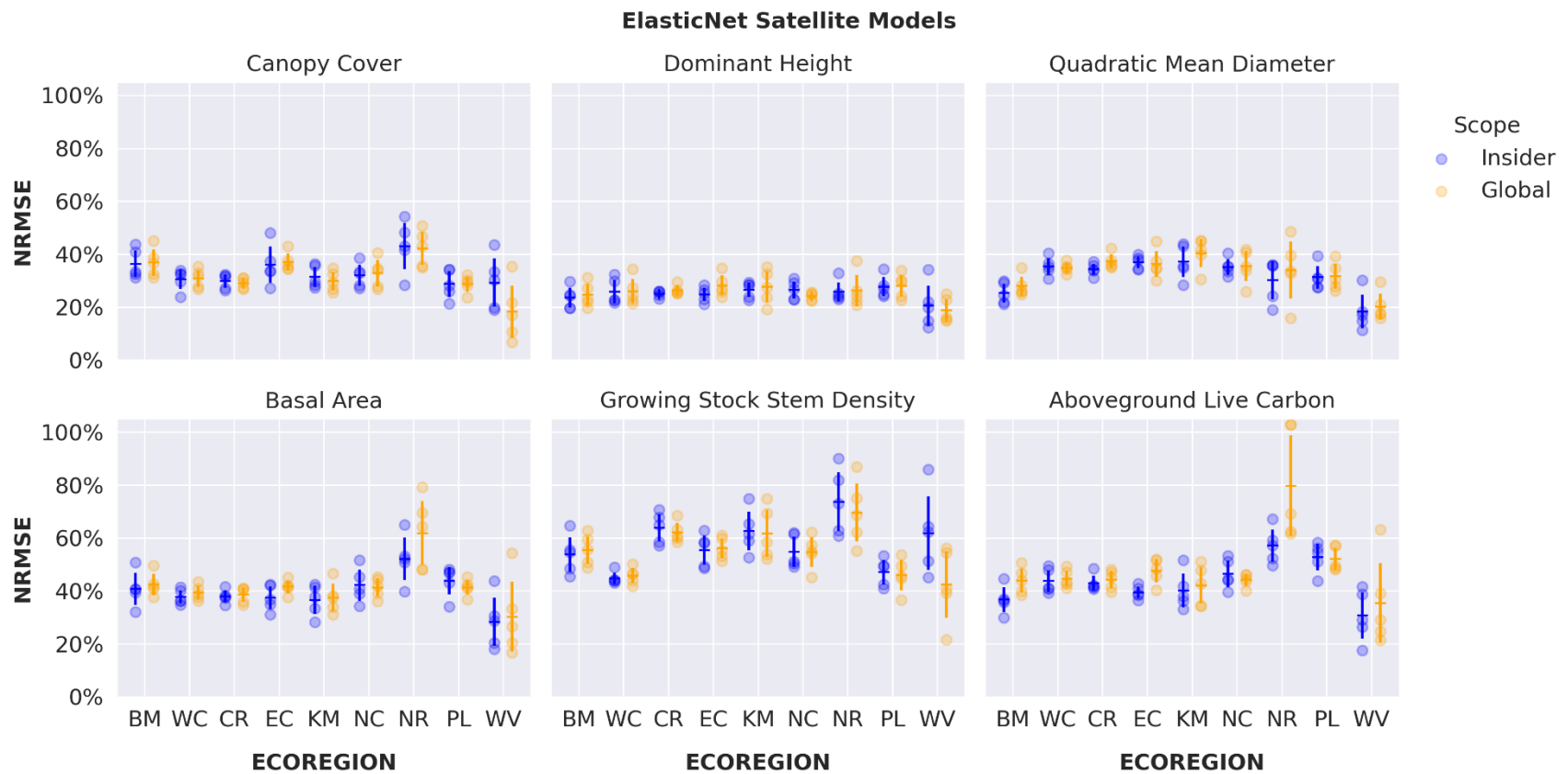


Figure S2.34. Performance (Normalized Root Mean Square Error, NRMSE) of ElasticNet-GLOBAL and ElasticNet-INSIDER satellite models across all ecoregions.

These graphs show the scores from five rounds of cross-validation for the ElasticNet Insider and Global models trained from satellite data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.

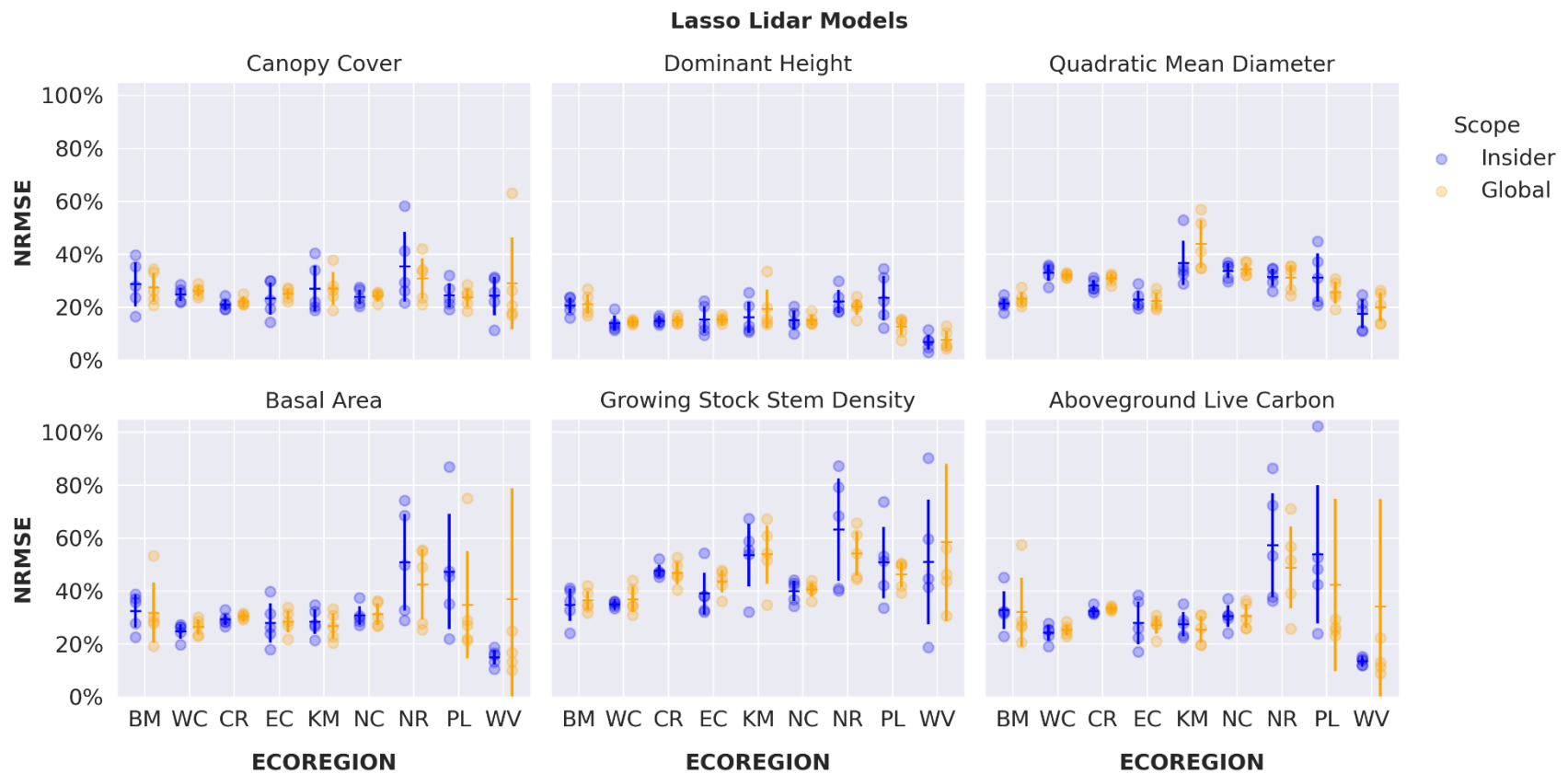


Figure S2.35. Performance (Normalized Root Mean Square Error, NRMSE) of LASSO-GLOBAL and LASSO-INSIDER lidar models across all ecoregions.

These graphs show the scores from five rounds of cross-validation for the Lasso Insider and Global models trained from lidar data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.

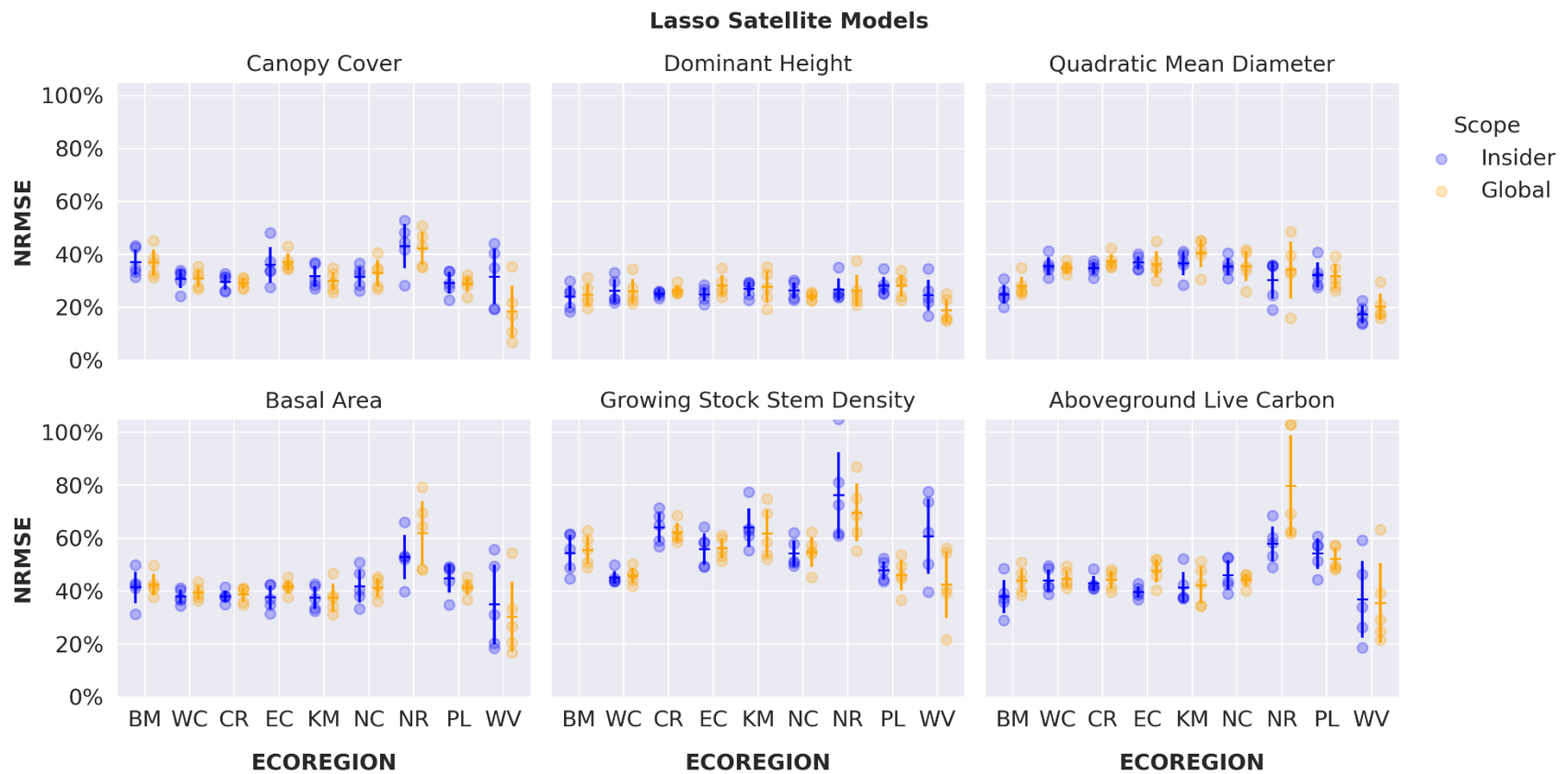


Figure S2.36. Performance (Normalized Root Mean Square Error, NRMSE) of LASSO-GLOBAL and LASSO-INSIDER satellite models across all ecoregions.

These graphs show the scores from five rounds of cross-validation for the Lasso Insider and Global models trained from satellite data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.

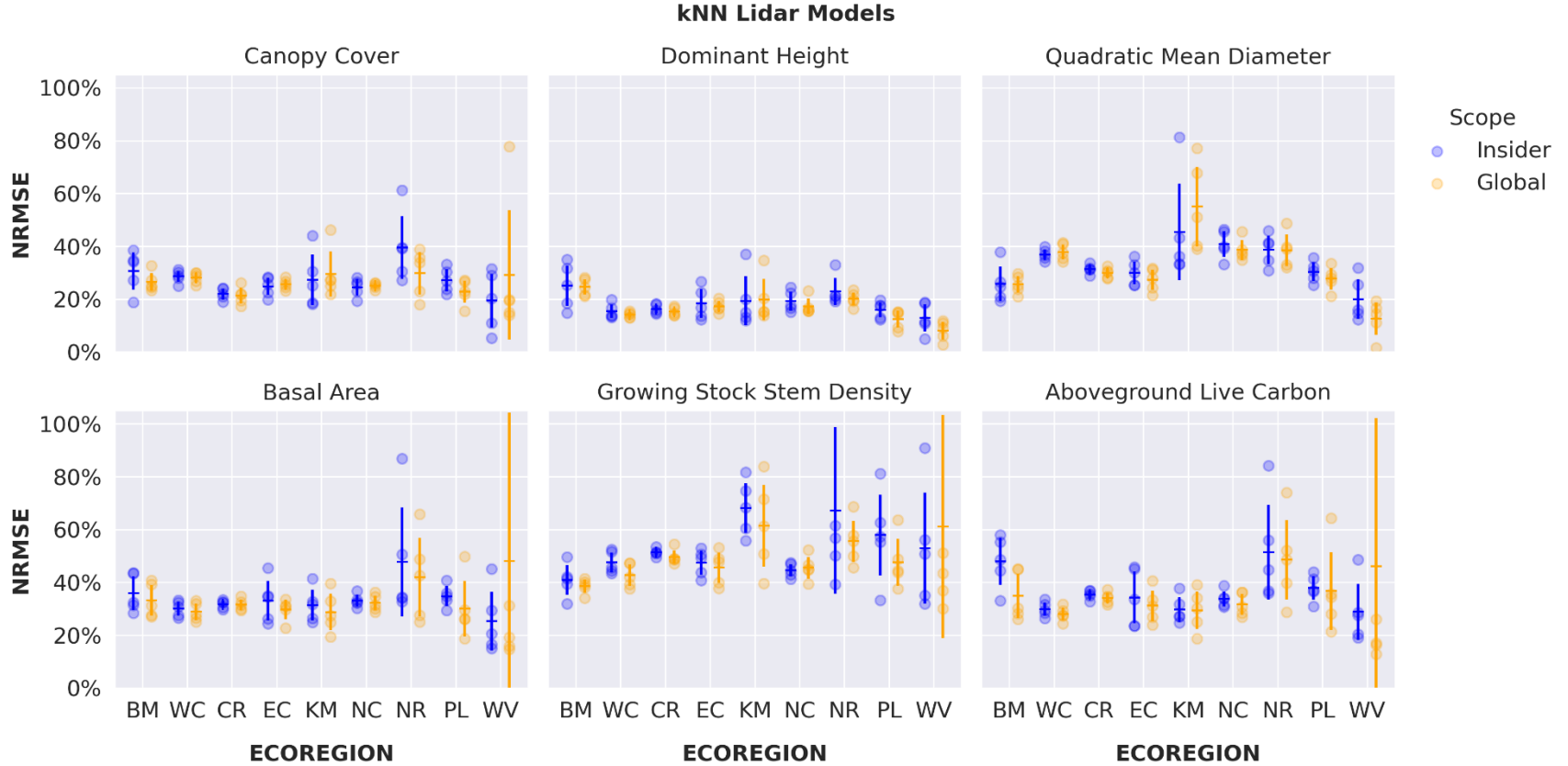


Figure S2.37. Performance (Normalized Root Mean Square Error, NRMSE) of KNN-GLOBAL and KNN-INSIDER lidar models across all ecoregions.

These graphs show the scores from five rounds of cross-validation for the kNN Insider and Global models trained from lidar data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.

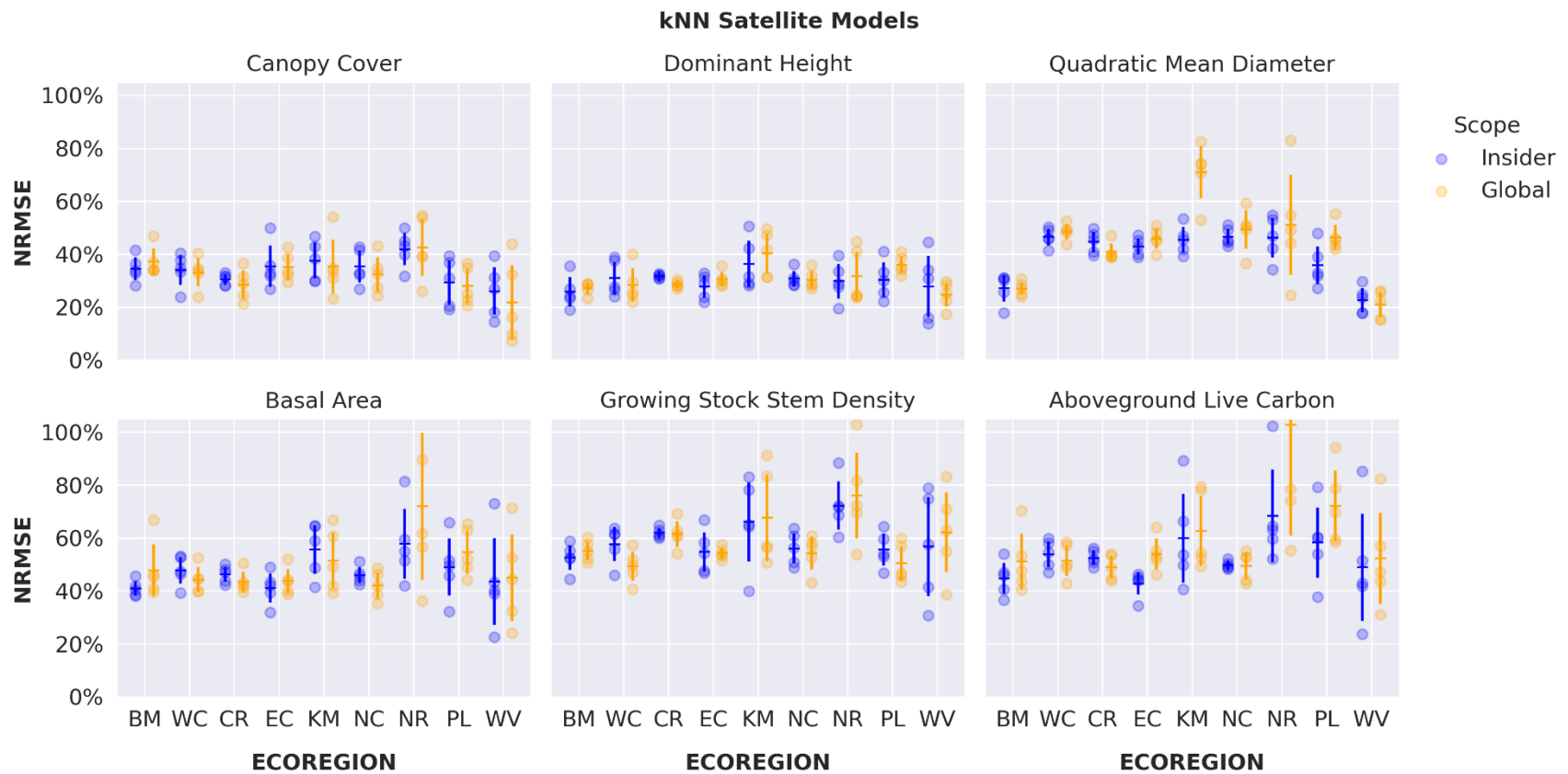


Figure S2.38. Performance (Normalized Root Mean Square Error, NRMSE) of KNN-GLOBAL and KNN-INSIDER satellite models across all ecoregions.

These graphs show the scores from five rounds of cross-validation for the kNN Insider and Global models trained from satellite data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.

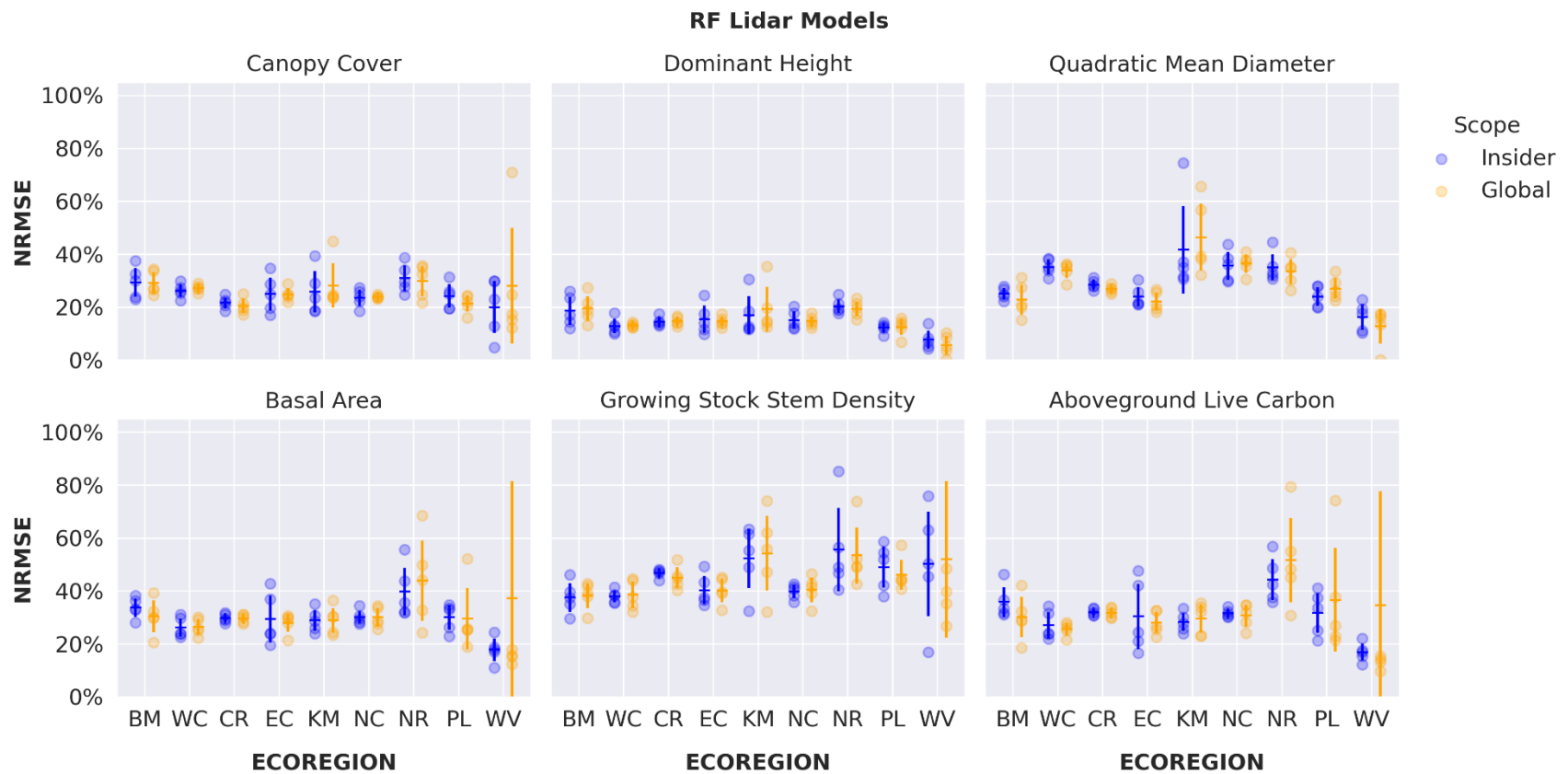


Figure S2.39. Performance (Normalized Root Mean Square Error, NRMSE) of RF-GLOBAL and RF-INSIDER lidar models across all ecoregions.

These graphs show the scores from five rounds of cross-validation for the RF Insider and Global models trained from lidar data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.

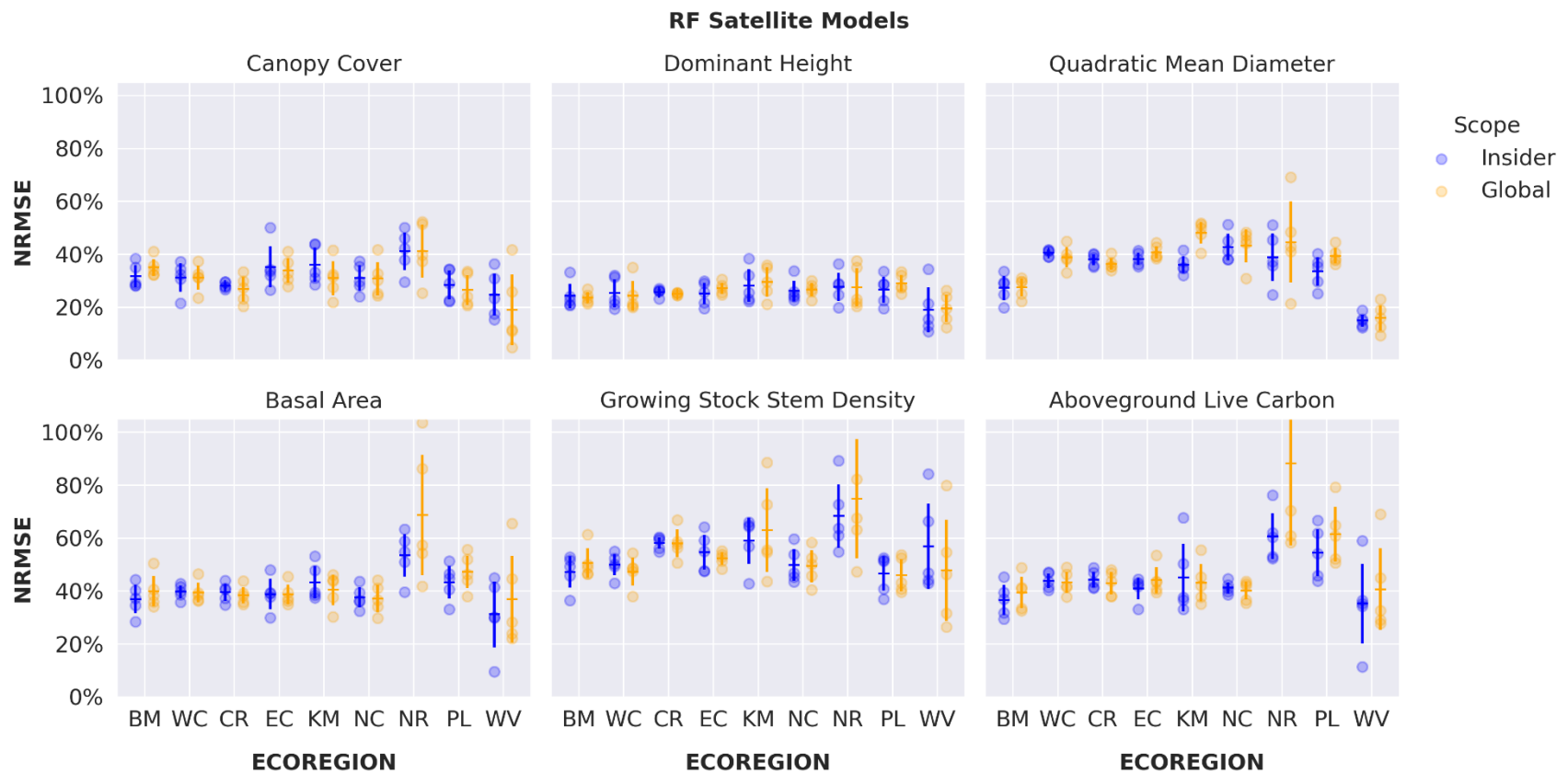


Figure S2.40. Performance (Normalized Root Mean Square Error, NRMSE) of RF-GLOBAL and RF-INSIDER satellite models across all ecoregions.

These graphs show the scores from five rounds of cross-validation for the RF Insider and Global models trained from satellite data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.

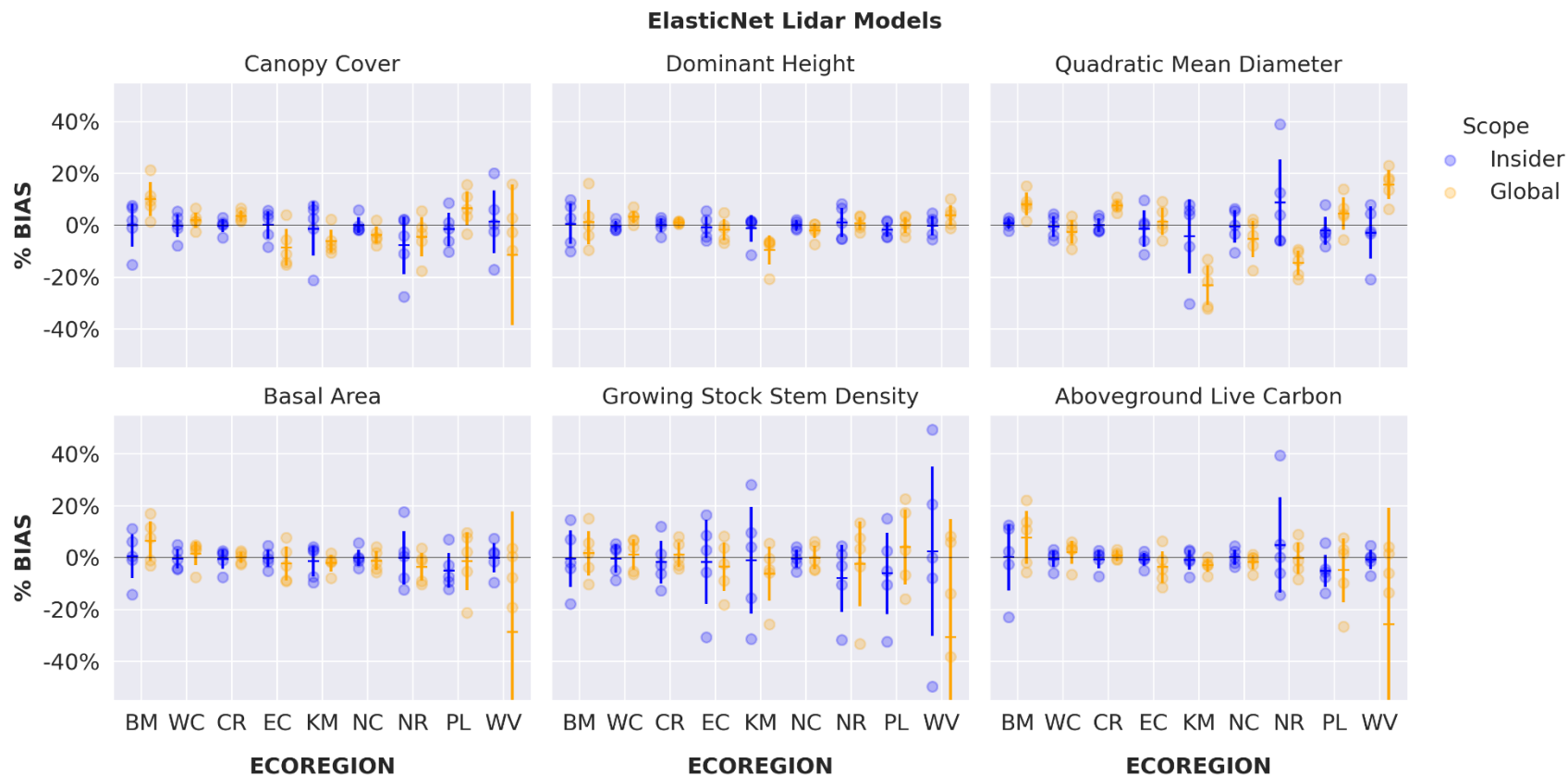


Figure S2.41. Performance (% bias) of ElasticNet-GLOBAL and ElasticNet-INSIDER lidar models across all ecoregions. *These graphs show the scores from five rounds of cross-validation for the ElasticNet Insider and Global models trained from lidar data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.*

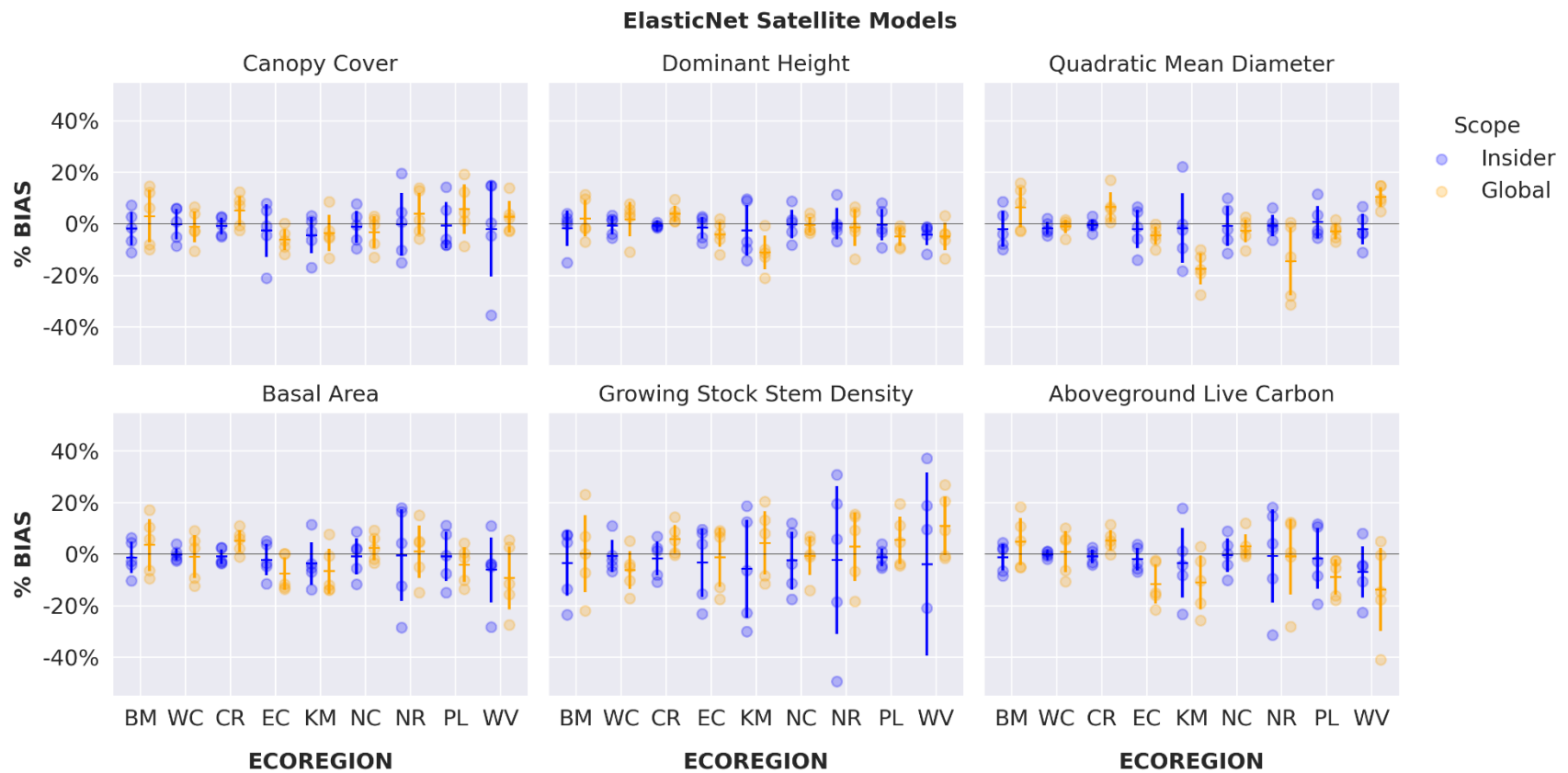


Figure S2.42. Performance (% bias) of ElasticNet-GLOBAL and ElasticNet-INSIDER satellite models across all ecoregions. These graphs show the scores from five rounds of cross-validation for the ElasticNet Insider and Global models trained from satellite data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.

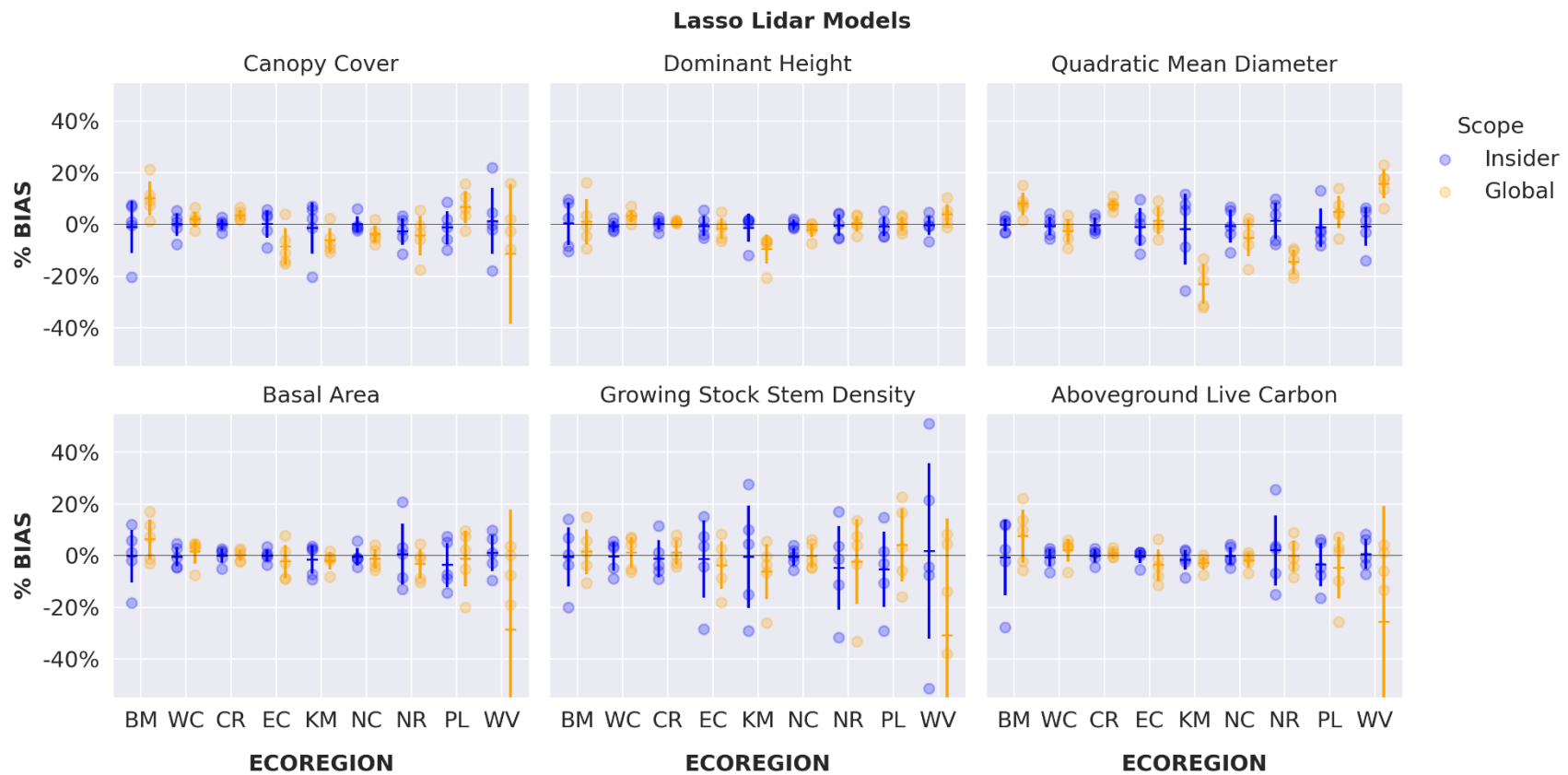


Figure S2.43. Performance (% bias) of LASSO-GLOBAL and LASSO-INSIDER lidar models across all ecoregions.

These graphs show the scores from five rounds of cross-validation for the Lasso Insider and Global models trained from lidar data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.

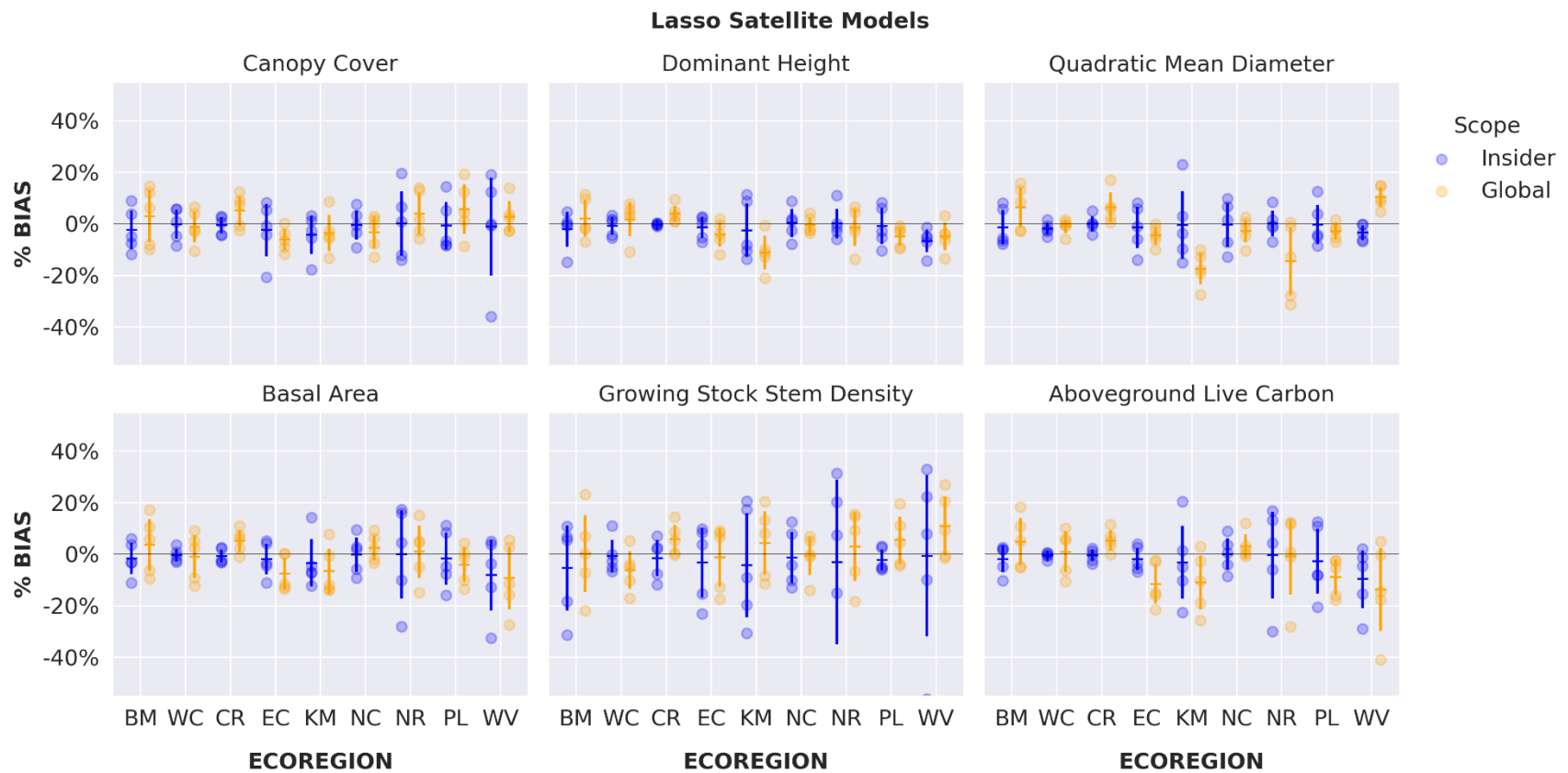


Figure S2.44. Performance (% bias) of LASSO-GLOBAL and LASSO-INSIDER satellite models across all ecoregions. These graphs show the scores from five rounds of cross-validation for the Lasso Insider and Global models trained from satellite data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.

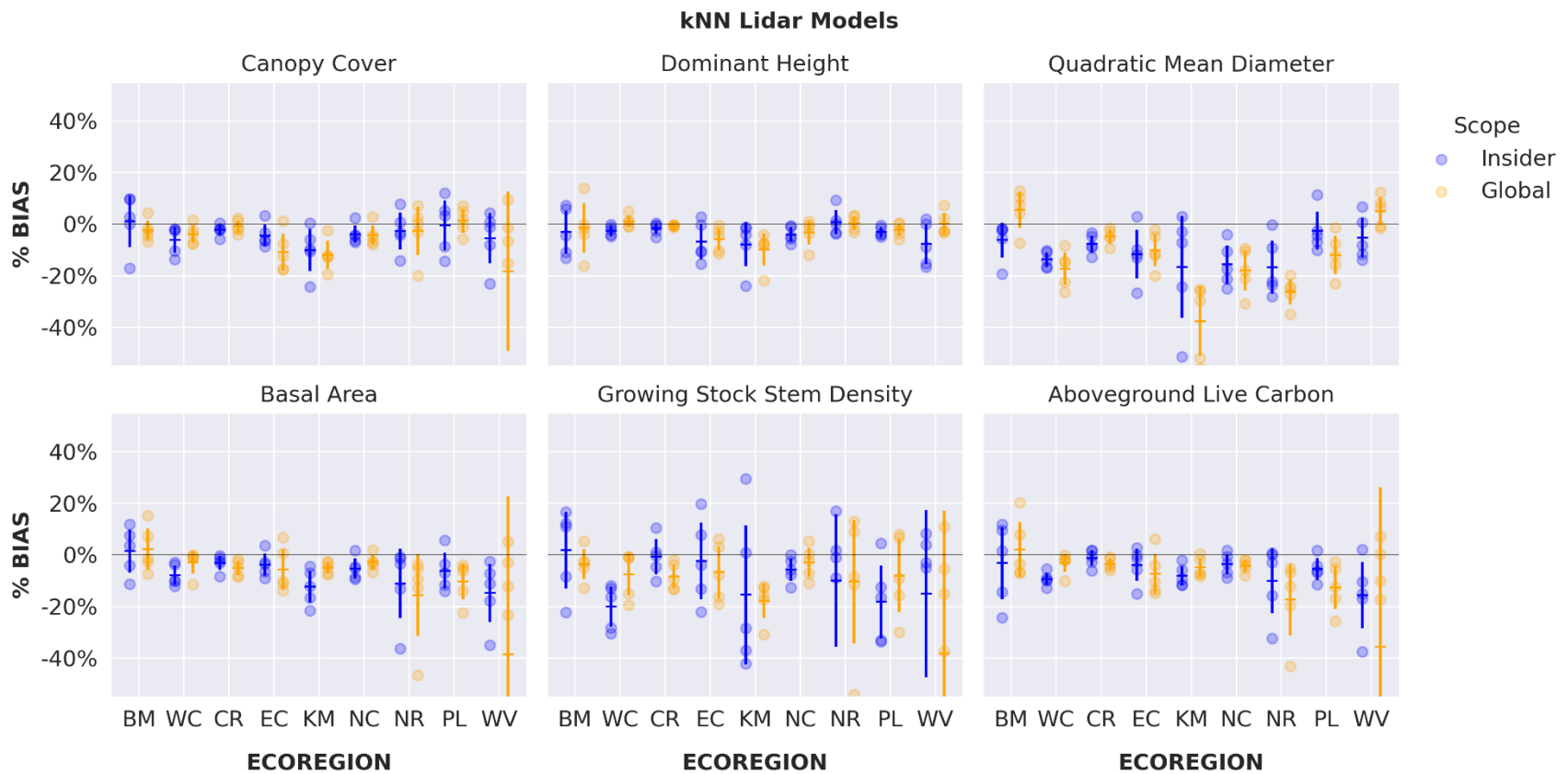


Figure S2.45. Performance (% bias) of KNN-GLOBAL and KNN-INSIDER lidar models across all ecoregions. These graphs show the scores from five rounds of cross-validation for the kNN Insider and Global models trained from lidar data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.

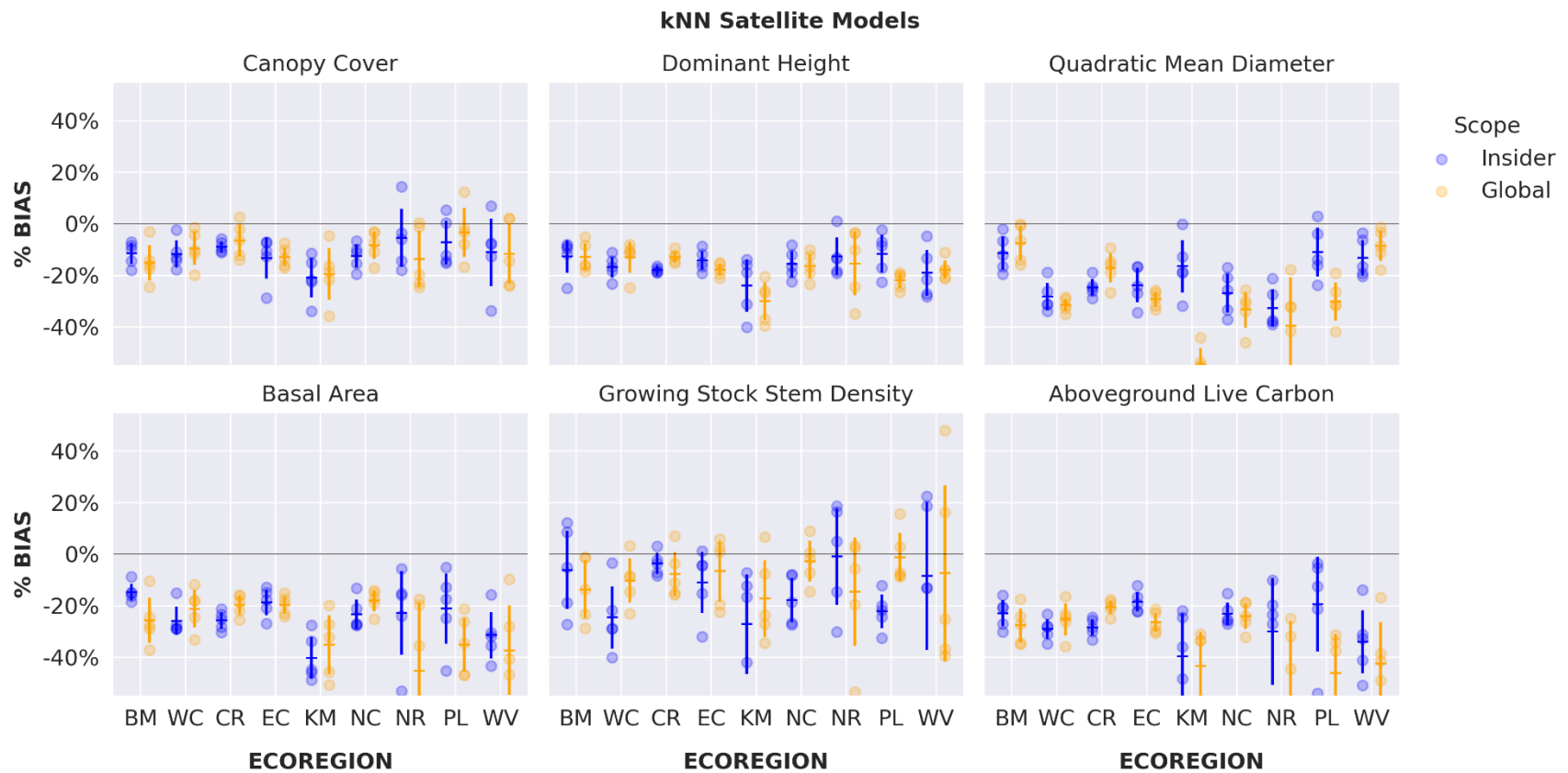


Figure S2.46. Performance (% bias) of KNN-GLOBAL and KNN-INSIDER satellite models across all ecoregions. These graphs show the scores from five rounds of cross-validation for the *kNN Insider* and *Global* models trained from satellite data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.

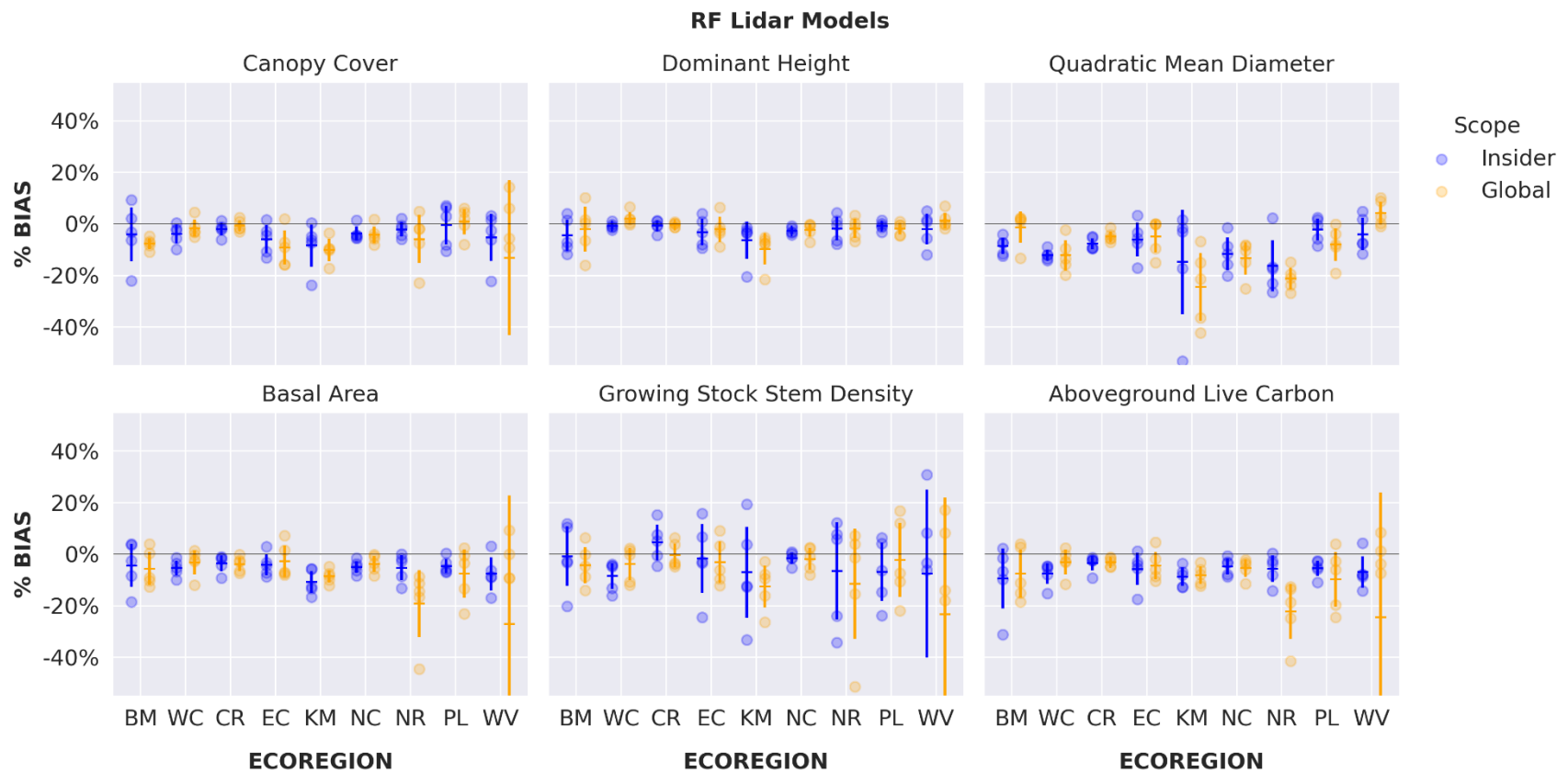


Figure S2.47. Performance (% bias) of RF-GLOBAL and RF-INSIDER lidar models across all ecoregions. These graphs show the scores from five rounds of cross-validation for the RF Insider and Global models trained from lidar data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.

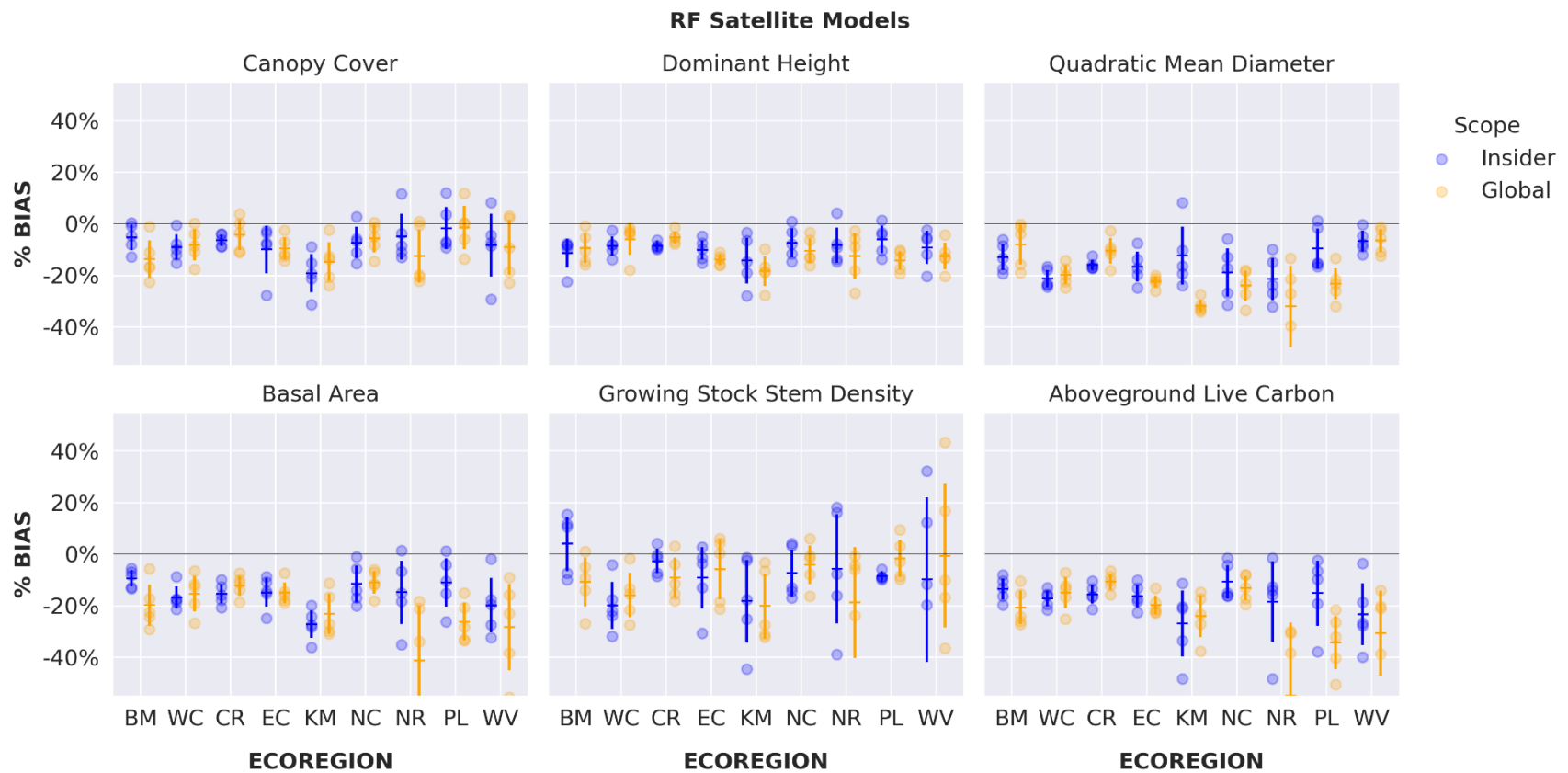


Figure S2.48. Performance (% bias) of RF-GLOBAL and RF-INSIDER satellite models across all ecoregions.
These graphs show the scores from five rounds of cross-validation for the RF Insider and Global models trained from satellite data and applied in each ecoregion. Lines running through the points indicate the mean plus/minus one standard deviation.

Chapter 3. A reproducible framework to refit Forest Vegetation

Simulator growth equations simultaneously across multiple ecoregions

3.1 INTRODUCTION

In the context of accelerating environmental change and the emergence of novel ecosystems, planning and implementing natural resource management and conservation will increasingly rely on predictive models and a shift away from long-standing reliance on assumptions of equilibrium; as systems become increasingly chaotic and novel, we expect predictability to suffer (McKelvey et al., 2021). Consistent with a rising emphasis on actionable science and research co-production, contemporary ecosystem modeling efforts need to adapt and be updated frequently, to be more inclusive in seeking out and addressing shifting user concerns and use cases as they arise, and to better quantify and communicate uncertainty and confidence in accessible ways (Cash et al., 2003; Kirchhoff et al., 2013; Gerber et al., 2020). It is also critical that the development of predictive models for ecosystem science and management be pursued in a way that allows them to be effectively documented and maintained over time, including distinct cultural and institutional contexts that may be downplayed or overlooked with the growing emphasis on user-centered design and engineering (Pearman and Cravens, 2022).

A long-standing debate has been waged over the properties and tradeoffs of “process-based” models that attempt to define fundamental models of ecophysiological processes and combine them into a structured representation of complex system dynamics and behavior with “empirical” models that are focused more on capturing statistical correlations between predictor variables and desirable outcomes with less concern over inference regarding the causal processes (Korzukhin et al., 1996). In practice, this distinction is increasingly seen as a false choice, as models that follow either purely process-based or purely empirical approaches are rare. Most models for forest ecosystems, particularly those that are widely used for natural resource planning and management, tend to incorporate ecophysiological principles where feasible following a “hybrid” approach, recognizing the challenges for users to access and employ data on critical ecophysiological drivers, (Korzukhin et al., 1996; Landsberg, 2003; Weiskittel et al., 2009; Adams

et al., 2013). As computational and statistical advances have been made, Bayesian approaches to analysis have become more accessible and effective, and these methods are being increasingly adopted to more intuitively quantify and communicate uncertainty in forecasting models using stochastic and probabilistic approaches (van Oijen, 2017; Wilson et al., 2019; Cailleret et al., 2020; Heilman et al., 2022).

The Forest Vegetation Simulator (FVS) is a widely used growth-and-yield model developed and maintained by the US Forest Service (USFS) (Crookston and Dixon, 2005; Dixon, 2020). FVS is commonly employed as a decision-support tool for forest planning and management, and as a research tool for forest science, policy development, and assessment. FVS is one of a small set of models that have been explicitly authorized for use in several forest carbon offset standards, and has subsequently seen wide use across the United States for this purpose. Examples of FVS applications include fire management (Fulé et al., 2004; Finney et al., 2007; Fried et al., 2017), economic analyses (Renner and Martin, 2002; Gilmore, 2003; Diaz et al., 2018), climate mitigation and adaptation potential (Bagdon and Huang, 2014; Gálvez et al., 2014; Tarancón et al., 2014; Diaz et al., 2015; Sánchez Meador et al., 2015; Hoover et al., 2021), habitat assessment (Maffei and Tandy, 2002; Zielinski et al., 2010), insect and disease dynamics (Atkins and Lundberg, 2002; Roberts, 2002; Russell et al., 2015), tradeoffs of ecosystem service provision with services under competing management scenarios (Schroder et al., 2016), and even auction approaches attempting to monetize ecosystem services (Tóth et al., 2010; Rabotyagov et al., 2013). In Oregon and Washington, FVS is the growth-and-yield model employed by state forestry and natural resource departments and federal agencies such as the USFS and Bureau of Land Management (BLM), influencing silvicultural decision-making and forest planning across more than 60% of Oregon forests and more than 50% of Washington forests.

Despite the fact that FVS developers have prescribed a framework for research applications of FVS to document their use of the software (Anderson et al., 2020), encouraged users to investigate the model's fit prior to adopting it for policy analysis or decision-support (Vandendriesche, 2010), and several illustrative examples of calibration and validation have been published (e.g., Shaw et al., 2006; Russell et al., 2013; Bagdon et al., 2021), many policy analyses and applications of FVS continue to employ FVS with default settings “out of the box”. A particular concern that has arisen in recent years are repeated findings in multiple regional variants of FVS that simulations run with default settings often show substantial and systematic biases to

overestimate forest growth (Canavan and Ramm, 2000; Froese and Robinson, 2007; Pokharel and Froese, 2008; Diaz et al., 2018; Herbert et al., 2023). Left unaddressed, systematic biases that overestimate growth and/or underestimate mortality have the potential to translate into highly misleading analyses used in policy and management decisions regarding timber volume and value, the potential for carbon sequestration to mitigate climate change, and forest resilience or adaptability to climate change.

Model fitting and calibration

FVS regional variants (see Figure 3.1) have traditionally been developed by fitting species-specific growth and mortality equations to remeasured plots within the intended footprint of the variant. Remeasured plot data have come from a variety of sources including silvicultural experiments and long-term research sites maintained primarily by USFS and in some cases based on timber industry research cooperatives. The proliferation of alternative model forms used for growth and mortality functions for different species and different regions without clear

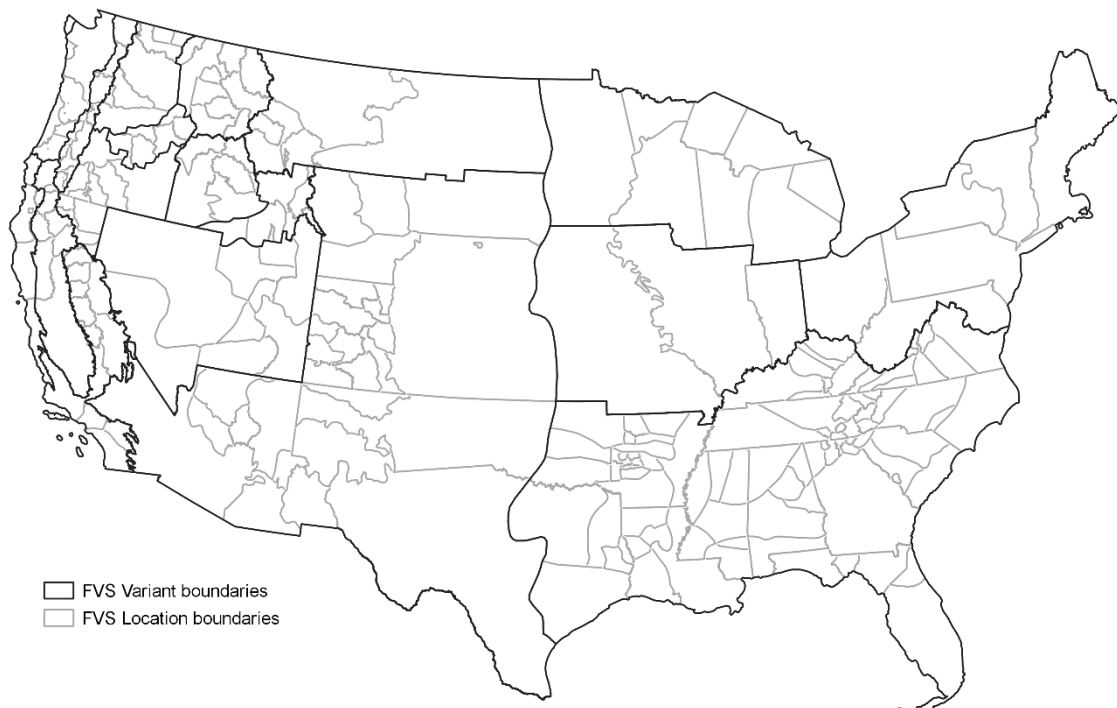


Figure 3.1. FVS Variants and Locations across CONUS

Across the contiguous United States, FVS software is released in 19 geographically distinct “Variants”. 15 of FVS’s total 19 Variants occur across 17 states in the US West. Sub-regional “FVS Locations” are delineated with light gray lines. Within each FVS Variant, growth, mortality, and allometric equations may be distinct for a tree species among FVS Locations.

justification or documentation can be seen as a form of “technical debt”, where a series of small tweaks and customizations to independent branches of the software accumulate over time and require a more time-consuming refactor and analysis to replace and maintain. Comments in FVS source code indicate that some growth and mortality routines were originally developed using observations exclusively from other regions from where they are being applied, and that many routines have not been revisited in several decades as larger and more robust datasets have become increasingly available.

The large tree diameter growth equation employed in the Pacific Northwest Coast (PN) Variant of FVS allows up to 18 non-intercept terms, though most individual species utilize only 11-14 (Keyser, 2017). Comments in FVS source code for diameter growth prediction in the PN Variant⁹ indicate coefficients for major species including Douglas-fir, Western redcedar, and Western hemlock were last updated in 1996, and that coefficients for some species (e.g., Ponderosa pine and Jeffrey pine) are copied from models fit exclusively to observations of these trees in other regions.

Software development and maintenance

FVS is a suite of regionally tailored open-source software programs, with source code primarily written in Fortran 77. While FVS source code was published on SourceForge for many years and recently migrated to GitHub (USFS FMSC, 2023), the development of FVS has generally been confined to a small team of USFS employees, with notable exceptions including development of regional variants in Ontario (Woods and Robinson, 2008) and in the Acadian region of the northern Maine and New Brunswick, Canada (Weiskittel et al., 2017). In light of the broad user base and significant impacts that the use of FVS shapes in public and private forests of North America, collaborative or community development following patterns demonstrated in a variety of other widely-used open-source software programs for scientific computing remains an alluring but unmet opportunity. The research reported in this dissertation chapter included direct consideration of three of the top five priorities identified recently by the Forest Carbon Modeling Group: increasing the diversity of forest conditions against which FVS is fit and calibrated for

⁹ <https://github.com/USDAForestService/ForestVegetationSimulator/blob/main/pn/dgf.f>

(#1); dynamically calibrating FVS against the FIA dataset (#2); and enabling more robust propagation of uncertainty (#5) (see Table S3.10).

A substantial set of model validation protocols for FVS were published more than ten years ago (Cawrse et al., 2010). However, subsequent efforts to recalibrate individual model components or regional variants of FVS are not consistently accompanied with published documentation (whether in the form of developer documentation, user guides, or scientific articles) demonstrating compliance with those protocols. As a result, many users and stakeholders of analyses completed using FVS software are left with relatively poor insights or intuition regarding how well the model should be expected to fit or reproduce the conditions or trends that may be found in the best available data or how well the model conforms to scientific principles of forest ecology and management sciences.

The limited implementation of unit testing and continuous integration systems is also a substantial hurdle that will need to be overcome to allow the potential for community support and maintenance of FVS software. There is a strong reliance on users to identify bugs that might otherwise be picked up through unit testing or functional testing. Test suites for unit or functional testing are essentially absent from the codebase. Although it is common for community-supported projects to require contributors to develop unit tests to correspond with new contributions of source code, a testing framework needs to be in place to allow this.

As the FVS team begins refitting some of the underlying equations for tree growth and mortality in regions across the western US (Michael Van Dyck, personal communication), it is my hope that the approach presented here offers useful steps towards the establishment of a robust and reproducible framework that will allow regional FVS models to be updated more frequently, consistently, and to be more easily maintainable as a steady stream of forest observations continue to emerge from the national network of the Forest Inventory & Analysis (FIA) program.

The FIA program as a robust and consistent data stream for calibrating FVS

The absence of shared datasets to benchmark models has not only translated into a lack of reporting of model performance as new versions of FVS software are released, but also limited the potential for developers to identify the strongest drivers of structural uncertainty or bias and contribute improvements to these models. The adoption of a benchmarking dataset and suite of

metrics and indicators used to evaluate model performance would offer significant benefits for model documentation and a proving ground to evaluate and justify community contributions.

Following the nationwide adoption of a consistent sampling framework and implementation beginning in 1999, measurements collected through the extensive network of FIA plots have rapidly accumulated and offered an unprecedented level of detail from which forest conditions and dynamics may be observed. Although the FIA sampling protocol was not primarily designed with growth-and-yield calibration in mind, its use in FVS has been increasingly facilitated and encouraged, including the release of FVS-ready data tables alongside raw FIA data (Shaw and Gagnon, 2020a).

The unbiased sample design, volume of data collected, and consistent formatting and ongoing public releases of FIA data provide a unique opportunity for data reflecting the full diversity of contemporary forest conditions and dynamics across US forests to be adopted as the primary source for model fitting, calibration, and benchmarking. Additional measurements are occasionally included in FIA samples, such as collection of tree cores for dendrochronological analysis and observation of additional soil attributes, increasing the potential for data-driven improvements to models of forest growth, mortality, and natural regeneration with the addition of down-scaled climate data. Although the FIA dataset may not provide the same degree of precision for tracking tree-level incremental growth and mortality as datasets collected from silvicultural trials and experimental/research forests due to the revisit frequency of FIA plots (10 years in Oregon and Washington, shifting to 5 years in California), the number of independent samples and diversity of forest and environmental conditions reflected in the FIA dataset suitable for growth-and-yield modeling is currently unmatched by any other program that releases data to the public.

Across coastal and western regions of Washington, Oregon, and California, FVS documentation reports sample sizes and data sources originally used to fit diameter growth models or “species relationships” (FVS Staff, 2022a, 2022b, 2023a), except for the Inland California and Southern Cascades (CA) FVS variant, which does not report sample sizes (FVS Staff, 2023b). The sample data used previously to train FVS models came from several surveys of National Forests, Bureau of Land Management forestlands in Oregon, and a survey by the Bureau of Indian Affairs of the Quinault Indian Reservation. The most recent source of inventory data in the Westside Cascades (WC) FVS variant is from 1987. The most recent source of inventory data in the Pacific Northwest Coast (PN) variant is the Quinault inventory from 1992. The dates of the inventory data

used in the Klamath Mountains (NC) FVS variant are not reported, but include samples from four National Forests, the Hoopa Indian Reservation, and Simpson Timber Company.

The FIA program provides a volume of data that outpaces previous datasets used to fit FVS, and offers numerous qualities that make it well-suited for benchmarking and refitting FVS models. In the last round of plot remeasurements alone, FIA observations within the scope of each FVS regional Variant has outpaced the sample size used by the FVS team for fitting existing FVS Variants for most of the tree species for which new models were fit in this study (see Table 3.1). The data from remeasured FIA plots have all been collected within the past 10 years, representing a recent snapshot of contemporary forest dynamics. FIA also provides a design-unbiased sample of forest conditions across public and private lands. In contrast, the data that were available at the time and used to fit most current FVS growth models is biased very strongly to National Forests, and on inventories that were all conducted more than thirty years ago. In the context of observations of declining diameter growth based on tree ring records analyzed across FIA plots across the southwestern USA (Klesse et al., 2018) and increasing mortality in western US forests (van Mantgem et al., 2009), it seems clear that more frequent model recalibration than has been historically the case for FVS models may be desirable to better capture contemporary forest growth and mortality expectations.

Recognizing and propagating uncertainty in models

FVS is generally used in a deterministic mode, with individual simulations run for each management scenario of interest. Error and uncertainty that are introduced through inventory data collection, imputation of data that are commonly missing from some inventory records (e.g., crown ratios and tree heights), the application of several regression equations to predict growth and mortality, and the use of biometric equations to derive estimates of crown area, canopy cover, canopy layering, fuel loading, as well as volume and biomass, are essentially ignored. In the case of FIA data, we can benefit from the fact that all the tree-level variables required by FVS to simulate growth and mortality (DBH, tree height, and crown ratio) are reported for all live trees. Explorations of the sensitivity of biometric equations used in FVS have indicated this component of structural model uncertainty alone may dominate the uncertainty in assessments of tradeoffs in timber and carbon volumes and values (Fischer et al., 2017).

Table 3.1. Samples sizes of trees used to fit current FVS models and from FIA data used for refitting models in this study

SPECIES	FVS REGIONAL VARIANTS								R5 + R6
	WC		PN		WC+PN		NC		
Common Name	FVS	FIA	FVS	FIA	FVS	FIA	FVS	FIA	FIA
Douglas-fir	17,250	15,649	10,098	10,155	27,348	25,804	6,502	3,783	44,004
Ponderosa pine	432	293	432	2	432	295	389	103	19,439
Western hemlock	5,008	6,288	3,931	3,571	8,939	9,859	NA	130	10,926
White fir	1,044	891	1,044	1	1,044	892	297	179	7,474
Pacific silver fir	3,878	4,114	622	373	4,500	4,487	NA	0	5,511
Mountain hemlock	3,019	1,836	3,019	54	3,019	1,890	NA	7	4,069
Western redcedar	1,354	1,567	1,171	751	2,525	2,318	NA	4	3,134
Canyon live oak	NA	8	NA	2	NA	10	NA	442	2,555
Tanoak	1	5	1	0	1	5	2,376	1,961	2,252
Red alder	125	527	1,369	1,189	1,494	1,716	NA	50	1,796
California black oak	NA	68	4	4	4	72	198	285	1,607
Pacific madrone	70	190	70	65	70	255	467	478	1,432
Coast redwood	0	0	0	0	0	0	*	1,405	1,405
Bigleaf maple	89	424	89	527	89	951	NA	100	1,205
Noble fir	1,555	976	1,555	20	1,555	996	NA	1	1,114

*Notes: This table displays the number of samples (individual trees) used in model fitting. Values reported for FVS PN and WC Variants indicate those used for fitting large tree diameter growth model. Documentation of the FVS NC Variant describes these as “samples used to fit species relationship” and may or may not represent diameter growth samples. “NA” indicates that a species is not represented in that FVS Variant. Shaded cells indicate samples are not limited to the geographic extent of a single FVS Variant. R5 + R6 indicates all samples collected from USFS Regions 5 and 6 (California, Oregon, and Washington). * The FVS NC Variant recognizes coast redwood models were incorporated in 2021, but does not document a sample size or source for inventory data.*

Particularly in the context of modeling forest dynamics under a changing climate, researchers will reasonably be concerned with how much uncertainty in climate-affected forecasts are related to model parameterization, and how much is related to climatic change. Research into a probabilistic forest modeling system in the Pacific Northwest led researchers to note markedly different predictions between deterministic and probabilistic models fit using the same model structure (Wilson et al., 2019). Recent research exploring the addition of climatic effects to FVS growth models and fitting models using diameter tape and increment measurements has also illustrated the utility of a Bayesian approach to quantify, partition, and propagate uncertainty (Heilman et al., 2022). Although I did not seek to perform a detailed sensitivity analysis or portioning of model uncertainty in the current study, it is worth pointing out the inherent power

and utility for Bayesian modeling to intuitively describe uncertainty in model parameters and the potential for capturing hierarchical relationships.

Learning across ecoregions

I suggest the persistence of the architecture of FVS following numerous independently-calibrated regional variants is an artifact of historical irregularity in data availability for model fitting regional models rather than an intention to treat forest dynamics in neighboring ecoregions as if they were independent processes. This is further supported by the presence of comments in FVS source code indicating that parameters for certain species' growth or mortality equations have been directly copied from other regions. The recognition that processes governing tree growth and mortality across an individual tree species' range are not independent leads directly to the conceptual framework offered by multi-level or hierarchical models. Multi-level or hierarchical models approach the issue of fitting parameters across multiple groups in a way that reflects their relatedness and represents an alternative to modeling approaches that ignore group differences ("fully pooled") or that treat groups as if they were completely independent ("unpooled") (see Figure 3.2),

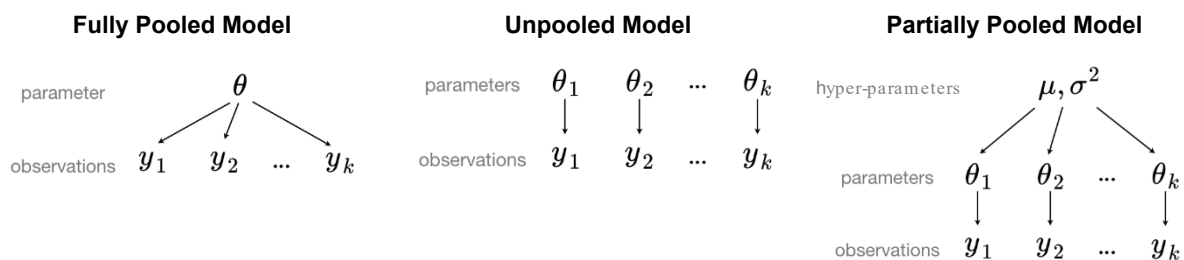


Figure 3.2. Hierarchical models as a partial pooling of parameters across related groups

For each of k groups, fully pooled models use a single set of parameters to estimate observations from all groups. Unpooled models use k independent parameters to estimate observations from each of the k groups. Partially pooled models define parameters for each of k groups as being drawn from a shared distribution among groups. Adapted from Chris Fonnesbeck, CC-BY-3.0¹⁰,

https://www.pymc.io/projects/examples/en/latest/case_studies/multilevel_modeling.html

The traditional pattern of FVS model calibration is best described in these terms as an unpooled approach. Parameters for a given equation within a regional Variant of FVS are traditionally fit to observations of trees and forests only from that region, with each regional

¹⁰ <http://creativecommons.org/licenses/by/3.0>

Variant being treated as a completely independent model. This is the most flexible approach to modeling, and may still be effective particularly for species and regions where many observations are available that also correspond to the conditions under which the model will be applied by users to make predictions. However, in regions where individual species are rare, this approach commonly leads to lumping together of species as “other softwoods” or “other hardwoods”.

Bayesian hierarchical models offer several potential advantages relative to the traditional FVS fitting approach of using unpooled regional parameters. A hierarchical model form can leverage a larger dataset gathered across multiple ecoregions, based on the assumption that there is meaningful information about the process that can be learned across ecoregions (i.e., that the process of diameter growth in each ecoregion is not independent of other ecoregions). The hierarchical approach allows the model to leverage or “borrow strength” from observations in ecoregions with numerous observations to inform parameters in ecoregions with few observations.

From a practical perspective of maintaining a software package, the potential to consolidate numerous regional variants into a system that offers broader geographic coverage for a larger number of species while also retaining localized parameters would significantly reduce the effort to refit and debug models over time. This utility would also be clearly beneficial to users interested in applying the model across diverse forest conditions without facing the current likelihood of discrete shifts in model performance when FVS is applied across ecoregion boundaries.

Primary research hypotheses explored

This chapter explores how the distribution of FVS-ready formats of FIA data can be leveraged to develop a reproducible open-source analytical pipeline to refit equations used in FVS in transparent and reproducible way as new FIA data are collected and released. I focused in particular on re-fitting the large tree diameter growth equation for several tree species across coastal Oregon, Washington, and California, but developed this workflow with the intent that it would be extensible to other equations, and across broad geographic extents.

In the interest of developing a multi-regional model-fitting approach that is easier to maintain, interpret, and eventually extend to include climatic drivers, I conducted four primary tests or sets of model comparisons in this study. First, I evaluated whether a single pair of variables for tree-level competitive status and stand-level competition could be adopted across all locations and ecoregions without compromising predictive performance. Then, I evaluated whether the full

set of terms currently employed in the Wykoff diameter growth model could be reduced without affecting predictive performance. Third, I investigated whether fitting the diameter growth model using a hierarchical approach alters predictive performance compared to the unpooled approach. And finally, I compared how the predictive performance of refitted models compares to estimates of diameter growth produced by the current public version of FVS.

I expected that the variety of FVS diameter model parameterizations among regions and species could be consolidated and simplified substantially without reducing model performance. I expected the adoption of a hierarchical model form would allow models to be fit using a more comprehensive set of observations without showing any substantial declines in predictive performance compared to the unpooled model form. I also expected that the newly refit diameter growth equation would show substantial improvement and a reduction in systematic biases observed in FVS models currently in use. If these three expectations are satisfied, I anticipate a much more straightforward approach to recalibrate FVS could be adopted on a recurring basis using FIA data. This would also enable the extension of these models to incorporate climatic drivers across multiple ecoregions.

3.2 METHODS

Terminology and levels of spatial representation

In this study, I considered drivers of forest growth across multiple spatial scales. The reasoning behind each of these methodological choices is included further below. The largest spatial scale at which drivers of diameter growth are modeled is the scale of an FVS Variant. In this study area, FVS Variants are similar in geographic extent to EPA Level III Ecoregions. Zooming in a bit further, FVS Locations divide the landscape of Variants into smaller areas based primarily on proximity to the nearest National Forest. The finest level of spatial representation occurs at the scale of FIA Plots and subplots. The terms “FVS Variant”, “FVS Location” and “FIA Plot” are used consistently throughout this study to refer to these geographic scales.

Data for model building and assessment

I focused effort on demonstrating this model-fitting workflow with attention to the most abundant tree species observed in coastal temperate forests from northern California up through

Washington state. This covered four regional Variants of FVS: Pacific Northwest Coast (PN), Westside Cascades (WC), Inland California and Southern Cascades (CA), and Klamath Mountains (NC). FIA data were gathered to cover the six most abundant tree species observed in each of these regions, resulting in 15 distinct tree species (see Table 3.). The geographic distribution of the FIA plots used in this study are shown in Figure 3.3.

I excluded FIA plots where no live trees were observed, as well as any plots within which a fire, geologic, or harvest disturbance was recorded between visits. I excluded these plots based on the expectation that FVS users might reasonably be expected to explicitly call for the inclusion of a fire or harvest event as part of a growth-and-yield simulation, and that geologic disturbances such as landslides and avalanches do not represent common occurrences a forest manager should expect to be factored into their growth-and-yield projections.

After eliminating plots subject to fire, harvest, or geologic disturbance, I intentionally retained plots where disturbances included damage or mortality due to insects, animals, disease, drought, and wind. I made this choice considering most forest modelers using FVS are unlikely to have information about these disturbances recorded in their forest inventory data, and that the effects of these types of disturbances may be difficult or impossible to model confidently in an FVS simulation. These types of disturbances commonly occur in managed forests and may substantially affect model bias (Glasby et al., 2019), although efforts to address these disturbances explicitly through additional inventory or modeling are very uncommon in practice. I therefore considered the inclusion of disturbances like these in training data to be an important strategy to avoid systematic overestimation of growth (and underestimation of mortality) that is traditionally described as inventory “falldown” (Bruce, 1977), or even as modeler “fall-up” when growth-and-yield models are calibrated using more “pristine” forest research plots.

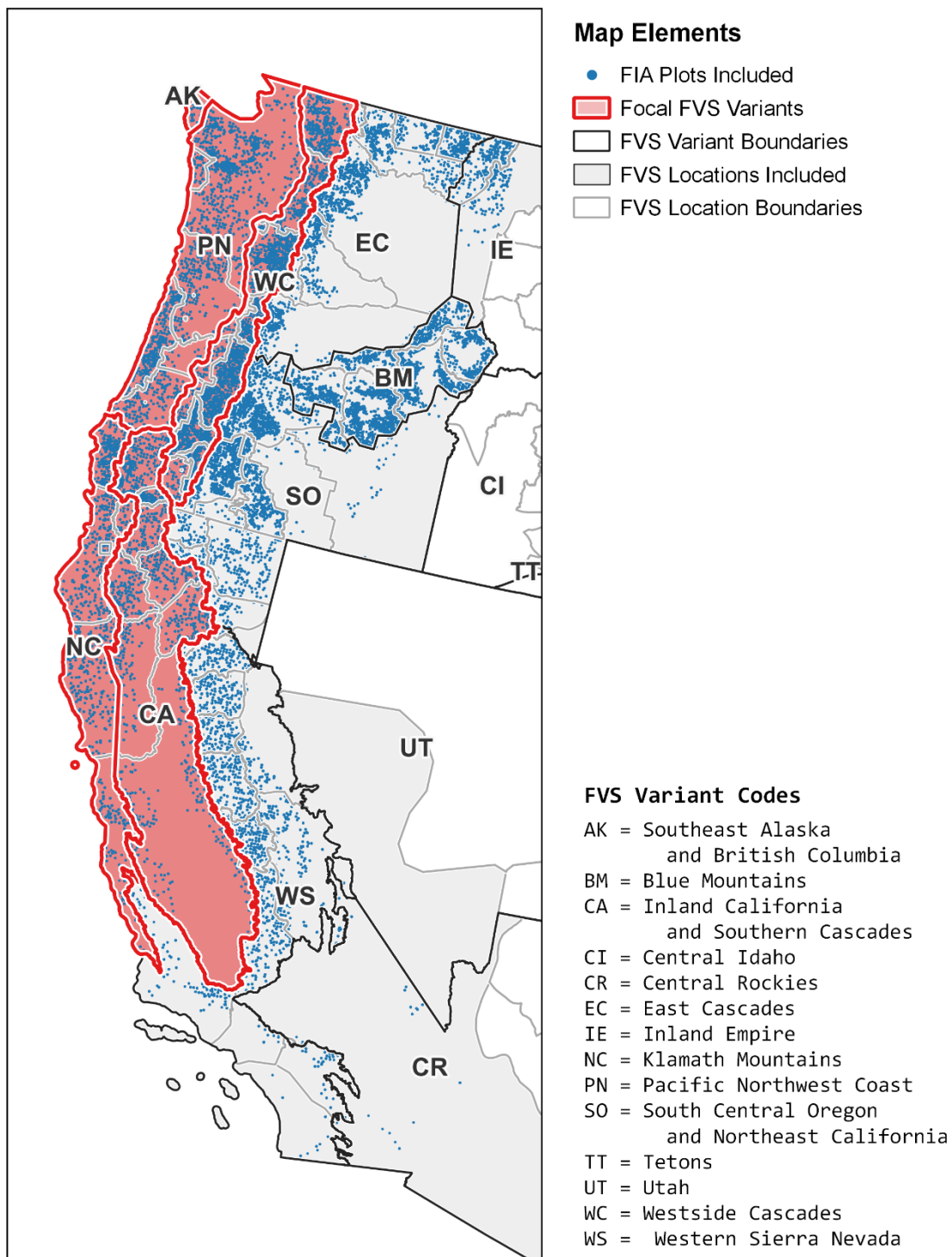


Figure 3.3. Layout of FVS Variants, Locations, and FIA Plot Data

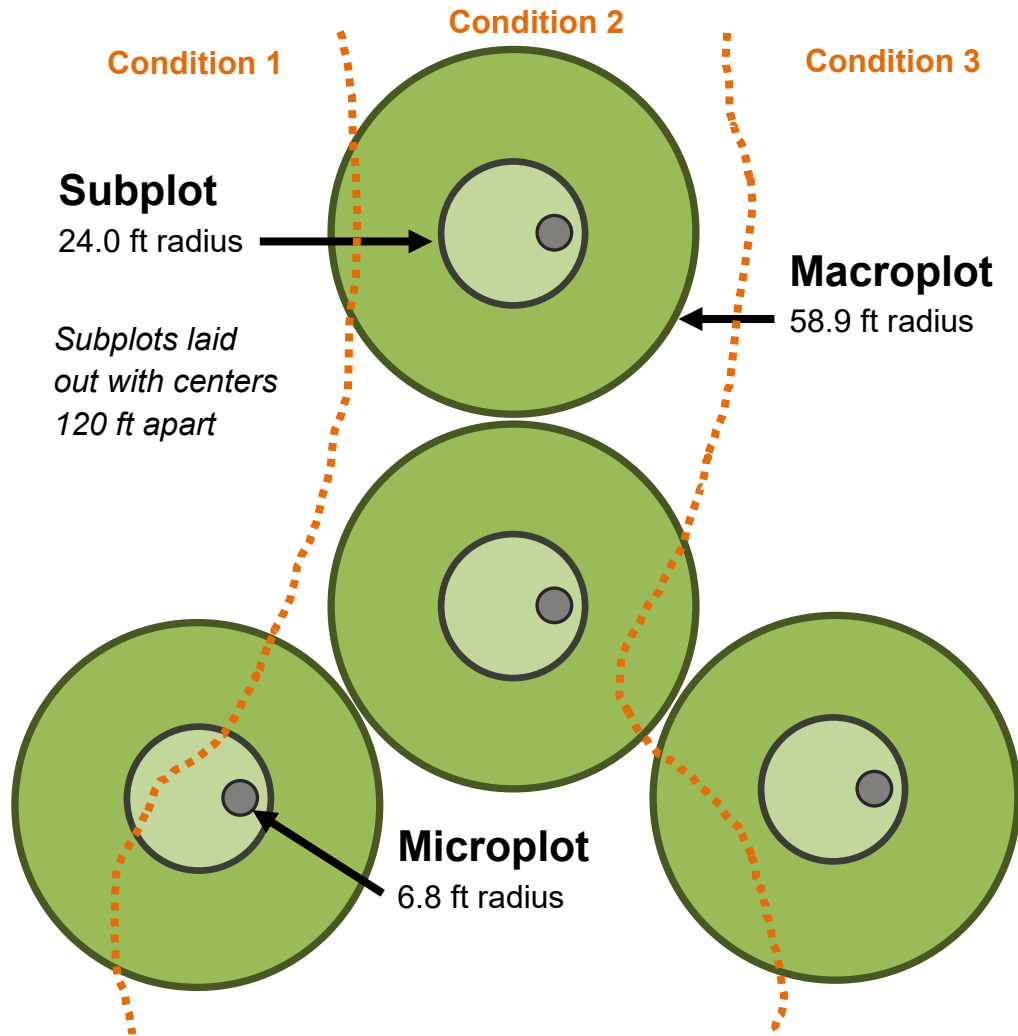


Figure 3.4. FIA plot layout with multiple conditions mapped across subplots
The delineation of three distinct forest conditions is shown across the four-subplot layout of each FIA plot.

FVS was originally conceived as a stand-level growth-and-yield model. Translating data from the FIA sampling design into FVS is commonly approached where the inventory data included for each “stand” follows one of three scopes: i) FIA “conditions”; ii) FIA plots, with multiple distinct subplots included; and iii) individual FIA subplots (Shaw, 2009; Shaw and Gagnon, 2020b, 2020a). FIA conditions (see Figure 3.4) are the most faithful representation of the stand concept (or whole FIA plots that occur within a single condition). However, in this study I opted to utilize the subplot as the inventory unit. Any subplot with remeasured live trees that wasn’t excluded due to disturbances was used, regardless of the forested proportion of the subplot or number of forested subplots within each FIA plot. I made the choice to use subplots as the spatial

scale for modeling growth anticipating the eventual application of FVS using inputs of inventory data derived from remote sensing, where correlative or generative models are employed to impute or synthesize inventory data in each pixel of a raster with pixel sizes commonly set to correspond to inventory plot sizes. This allows for forest edges and ecotones to be reflected in the design of the training data, but also comes with an expected cost of increased variation.

Finally, the last filter of the data involved limiting observations to those of trees recorded ten years apart. This is due in part to the intended fitting and application of the diameter growth model to predict five-year diameter increments, and to avoid complications involved with fitting a fully-annualized growth model or extrapolating the small proportion of measurements collected off the intended 10-year remeasurement cycle to an assumed 10-year-equivalent.

FIA data for each remeasured subplot was then prepared and augmented with calculations of tree-level attributes required as covariates for candidate diameter growth models, such as basal area of larger trees and relative height, as well as “stand”-level variables such as total live tree basal area and crown competition factor. Total heights for trees without field-measured heights were imputed using height-diameter regression equations fit to trees observed within each FVS Location (smaller geographic areas within each FVS Variant roughly corresponding to the nearest US National Forest). A hierarchical model using a Chapman-Richards functional form was fit using up to 5,000 samples of each species across all locations and used to impute heights for trees without recorded heights. More detail on the height imputation modeling is provided in Supplementary Materials.

Data from all repeatedly measured trees on eligible subplots were gathered and pre-formatted for model training and validation. Although these modeling efforts and interpretation were focused on a subset of this region, I gathered data from all eligible FIA plots across California, Oregon, and Washington. This included trees observed outside the four focal ecoregions or FVS variants. This choice was intended to help illustrate the potential for fitting a hierarchical version of the FVS diameter growth model capable of borrowing strength across ecoregions (described in greater detail further below).

Trees with recorded DBH decreases greater than the FIA program’s Measurement Quality Objective for DBH measurement error (0.1 inches of measurement error tolerated per 20 inches of DBH) were excluded from the analysis. For the fifteen tree species for which diameter growth models were re-fit, the number of individual trees available for model-fitting ranged from just over

Table 3.2. Repeatedly measured trees used for model training and evaluation

SPECIES	FVS REGIONAL VARIANTS									
Common Name	WC	PN	SO	EC	BM	CA	NC	WS	IE	Total
Douglas-fir	15,649	10,155	303	5,614	3,318	3,490	3,783	889	803	44,004
Ponderosa pine	293	2	8,315	2,099	6,057	1,023	103	1,078	469	19,439
Western hemlock	6,288	3,571	20	699	--	30	130	--	188	10,926
White fir	891	1	2,706	4	329	1,267	179	2,097	--	7,474
Pacific silver fir	4,114	373	88	932	4	--	--	--	--	5,511
Mountain hemlock	1,836	54	1,022	690	8	83	7	360	9	4,069
Western redcedar	1,567	751	--	338	--	9	4	--	465	3,134
Canyon live oak	8	2	--	--	--	1,470	442	633	--	2,555
Tanoak	5	--	--	--	--	198	1,961	88	--	2,252
Red alder	527	1,189	--	13	5	12	50	--	--	1,796
California black oak	68	4	37	--	--	637	285	576	--	1,607
Pacific madrone	190	65	--	--	--	688	478	11	--	1,432
Coast redwood	--	--	--	--	--	--	1,405	--	--	1,405
Bigleaf maple	424	527	--	30	--	98	100	26	--	1,205
Noble fir	976	20	10	103	1	2	1	--	1	1,114

Notes: The six most abundant species in each of the four focal ecoregions are shown with tinted cells. Each cell displays the count of unique live trees with repeated observations of DBH available for modeling.

1,000 to as high as 44,000 (see Table 3.). For Ponderosa pine and White fir, the number of samples from FVS variants outside the four focal ecoregions where I refitted models were larger than the sample size within the four focal regions.

The Wykoff process model for diameter growth

The primary conceptual foundation for FVS diameter growth modeling across western US variants can be found in an equation proposed by Wykoff (1990), which models individual tree basal area increment as a function of the tree's starting diameter, site-level biophysical variables, and indicators of tree- and stand-level competition. To be more precise, many of FVS's regional variants model the natural logarithm of the periodic change in squared inside-bark tree diameter (*dds*) following this general functional form:

$$\ln(dds_t) = SIZE_t + SITE + COMP_t \quad (3.1)$$

where SIZE, SITE, and COMP are determined by the following equations. The effect of SIZE was based on the tree's starting diameter at breast height (*dbh*):

$$SIZE_t = b_1 \ln(dbh_t) + b_2 dbh_t^2 \quad (3.2)$$

SITE was calculated from one or more intercept terms that reflect a forest's location and/or habitat type, insolation or exposure derived from a site's slope (*SL*) and aspect (*ASP*), and elevation (*EL*):

$$SITE = b_0 + b_3SL[\sin(ASP)] + b_4SL[\cos(ASP)] + b_5SL + b_6SL^2 + b_7EL + b_8EL^2 \quad (3.3)$$

Competitive effects (COMP) were determined from a tree's live crown ratio (*CR*), a stand-level indicator of total competition known as crown competition factor (*CCF*) and the stand basal area in live trees larger than the tree being modeled (*BAL*):

$$COMP_t = b_9CR_t + b_{10}CR_t^2 + b_{11}CCF_t + b_{12}BAL_t \quad (3.4)$$

These models are generally deployed in FVS to predict diameter growth at five- or ten-year timesteps. In this study, I adopted a five-year prediction interval.

Interpolating predictor variables between field observations

I fitted the diameter growth model as a recursive function, iteratively predicting inside-bark *dds* at each timestep, which was translated into an estimate of outside bark diameter growth and used to calculate DBH at the end of the timestep. The predicted DBH at the end of each timestep was used as an input for SIZE effects in subsequent timesteps. This series of conversions is provided below.

I first utilized the following equation to calculate diameter inside bark (*DIB_t*) from diameter outside bark (*DBH_t*) at the start of the growth period:

$$DIB_t = a_0 + a_1DBH_t^{a_2} \quad (3.5)$$

where a_0 , a_1 , and a_2 are bark ratio coefficients defined in the documentation of each regional FVS Variant. After calculating $\ln(dds_t)$ following Equation 3.1 and exponentiating it, the change in squared inside-bark diameter dds_t and estimate of diameter inside bark at the end of the five-year timestep were calculated as:

$$DIB_{t+5} = \sqrt{DIB_t^2 + dds_t} \quad (3.6)$$

From the estimated diameter inside bark at the end of the five-year timestep, diameter outside bark at the end of the timestep was calculated as:

$$DBH_{t+5} = \left(\frac{DIB_{t+5} - a_0}{a_1} \right)^{1/a_2} \quad (3.7)$$

With estimates of DIB and DBH at both the beginning and ending of each time-step, inside-bark and outside-bark diameter growth were calculated as the difference between DIB or DBH at the start and end of the timestep. Incremental radial growth was estimated as half of incremental diameter growth.

Following examples demonstrated by Cao (2000, 2004) and Weiskittel et al., (2007), I utilized linear interpolation to estimate intermediate values for covariates that evolve over time at timesteps where no field measurements were made. Although these earlier examples utilized this approach to move to annualized versions of growth and mortality equations from periodic observations, I employed the same general approach to estimate variables included in COMP fixed effects such as crown ratio, and tree- and stand-level competition variables at the five-year timestep. This allowed the growth model to be fit to individual trees without needing to carry forward all trees simultaneously to estimate changes in plot- or stand-level stocking variables and avoids the need to employ another regression model to predict changes in crown ratio at each timestep.

Mixed effects modeling

For each tree species within a particular FVS variant, ~12-15 variables are commonly used to predict diameter growth. Several alternative covariates are employed for tree- and stand-level competitive effects across FVS variants, and Site Index is occasionally incorporated as an additional covariate in the estimate of SITE effects. The covariates used within each FVS Variant differ among tree species, and the covariates used for a single species also commonly differ across FVS variants.

In addition to non-intercept terms used to describe fixed effects of tree SIZE, SITE, and COMP, I extended the Wykoff model form to add random effects at the levels of FVS Location and FIA plot. To be clear, while I fitted models to trees with competition indicators based only on trees within a single FIA subplot, I assigned a random effect at the scale of the FIA plot, which covers multiple FIA subplots. Random effects for FVS Location and FIA plot were fit using a sum-to-zero constraint to avoid non-identifiability (solution #2 identified by Ogle and Barber, 2020).

Simultaneously fitting these models across data from Washington, Oregon, and California, which span multiple FVS variants, I generated distinct intercept (b_0) and covariate coefficient (b_1)

- b_{13}) parameters for each FVS variant. The intercept terms (b_0), along with random effects terms ε_{loc} and ε_{plot} capture variation unexplained by the covariate coefficient parameters ($b_1 - b_{13}$) at the spatial scales of FVS Variants, FVS Locations, and FIA plots, respectively.

Candidate models evaluated

There are two general model forms that I evaluated, which I refer to as “FULL” and “SIMPLE” Wykoff models. The FULL models incorporated 13 non-intercept terms, with the addition of a term for site productivity (MAI) added to the fixed SITE effects described above. That is, the form I investigated for the “Full” model was:

$$\ln(dds_t) \sim N(SIZE_t + SITE + COMP_t + \varepsilon_{loc} + \varepsilon_{plot}, \varepsilon_{wyk}) \quad (3.8)$$

$$SIZE_t = b_1 \ln(dbh_t) + b_2 dbh_t^2 \quad (3.2)$$

$$SITE = b_0 + b_3 MAI + b_4 SL[\sin(ASP)] + b_5 SL[\cos(ASP)] + b_6 SL + b_7 SL^2 + b_8 EL + b_9 EL^2 \quad (3.9)$$

$$COMP_t = b_{10} CR_t + b_{11} CR_t^2 + b_{12} STAND_COMP_t + b_{13} TREE_COMP_t \quad (3.10)$$

Where: ε_{loc} and ε_{plot} are random effects at the scale of the FVS location and FIA plot, respectively; ε_{wyk} is the standard deviation of process error in the Wykoff model (Eq. 3.1), which is assumed to be normally distributed; *MAI* is potential Mean Annual Increment of cubic volume; *STAND_COMP* is a placeholder for one of three alternative stand-level competition variables, and *TREE_COMP* is a placeholder for one of three alternative tree-level competition variables. The model form employed in most FVS variants utilizes Site Index instead of MAI. Due to the wide variation in site species and base ages used in the FIA program to calculate site index across these regions, I calculated MAI from Site Index following Hanson et al. (2002) to provide a more consistent representation of site productivity across the study area than would be possible using Site Index values based on multiple site species and base ages.

For the “simplified” form of the Wykoff model, I dropped terms b_4 , b_5 , b_7 , b_9 , and b_{11} from the “full” model, removing fixed effects intended to capture topographic exposure and quadratic terms for *SL*, *EL*, and *CR*.

Although publicly-available versions of FVS occasionally utilize more than one variable for tree-level and/or stand-level competition effects, I opted to limit model selection to permit only

one variable for each of these fixed effects to improve model interpretability. For tree-level competition variables, I considered models using one of three alternative covariates. The first two options of covariates to indicate tree-level competitive status were basal area of trees larger than the modeled tree (BAL) and the height of the target tree in relation to the height of dominant trees (RELHT), which FVS caps at a value of 1.5. A third option to describe tree-level competition term operates effectively as an interaction with tree size: the basal area of larger trees divided by the natural log of the target tree's DBH plus one (i.e., $BAL / \ln(DBH + 1)$), referred to in the remainder of this study using the code BALLNDBH. For stand-level competition variables, I considered models using one of three alternative covariates: total basal area of live trees (BA), the natural log of BA (LNBA), or Crown Competition Factor (CCF) as defined by Krajicek et al. (1961) and calculated at the scale of the subplot using existing FVS equations for each species and variant.

To identify the various candidate models in this study, I refer to the combinations of model form (full or simplified), to the tree-level competition variable, and the stand-level competition variable: FORM-TREE_COMP-STAND_COMP. For example, SIMPLE-BAL-BA model refers to the simplified version of the Wyckoff model fitted using BAL for the tree-level competition variable and BA for the stand-level competition variable. For each of the full and simplified models, there are nine combinations of tree- and stand-level competition variables to choose from, leading to a total of eighteen models (nine FULL and nine SIMPLE) that will be compared.

State-space modeling using diameter and increment observations measured with error

In most of the western USA, FIA plots have been revisited on a 10-year target frequency. For the vast majority of trees on remeasured FIA plots, diameter growth can only be inferred from the change in DBH between two distinct field measurements. In addition to DBH measurements collected using a diameter tape, a subset of trees also have radial growth measured using increment cores, which are reported in terms of five- and ten-year observed radial increments. It is important to note that these increment cores are not collected for or subject to precise dendrochronology analyses in a lab setting, and are not expected to be as precise as similar studies which have employed tree ring records in addition to inventory diameter measurements.

Employing estimates of measurement error based on Measurement Quality Objectives defined by the FIA program, I incorporated two sources of observation error using a Bayesian errors-in-variables approach. Combined with the recursive use of predicted diameter growth as an

input to subsequent predictions, the modeling approach adopted here employed a Bayesian State-Space framework that is conceptually similar to previous work by Clark et al. (2007), Schliep et al. (2014), and Heilman et al. (2022). An overview of the Bayesian State-Space Model is illustrated in Figure 3.5.

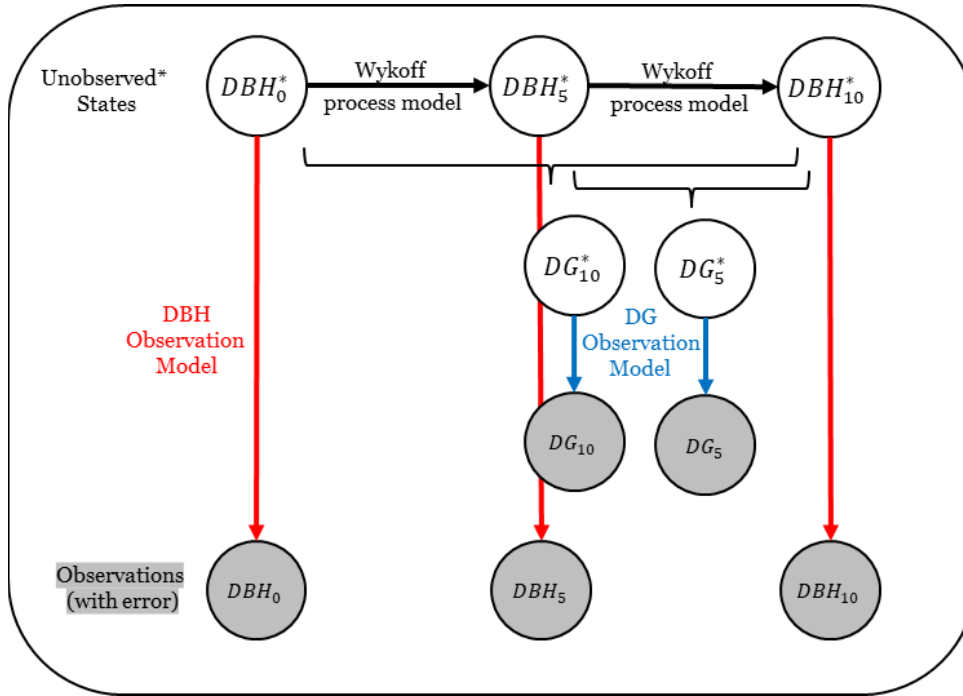


Figure 3.5. Illustration of the Bayesian State-Space model

The evolution of the hidden state (DBH) is governed by the Wykoff process model. The predicted future states of DBH are conditioned on FIA measurements using the DBH observation model. The predicted incremental growth from years 0-10 and from years 5-10 is conditioned on FIA measurements of radial increment on the subset of trees where this data was collected using the DG observation model.

The State-Space modeling approach utilizes two distinct types of models: a process model that describes a “true” but not fully observed state, and an observation model that relates the observations measured with error to the true state. The true tree diameter at breast height at any given time t , is denoted as DBH_t^* . I defined DBH_t^* to be conditionally dependent on measurements using a normal probability distribution with mean DBH_t and standard deviation of $0.01 \times DBH_t$. I describe this as the DBH observation model:

$$DBH_t^* \sim N(DBH_t, 0.01 \times DBH_t) \quad (3.11)$$

The standard deviation of the DBH observation model reflects increasing measurement error as tree size increases, and was estimated from reports of observed variation in DBH

measurements reported from quality assessment efforts for FIA plots in the Pacific Northwest that indicate the FIA Measurement Quality Objective (MQO) for DBH measurement error (USFS Forest Inventory and Analysis, 2023) which calls for 95% of measurements to be within 0.1 inches of the real DBH for each 20 inches of DBH (e.g., a tolerance of 0.2” for trees with 20” < DBH < 40”) may not be satisfied in this region (Melson et al., 2002).

An analysis to estimate DBH measurement error from the trends reported in Melson et al. (2002) error using FIA data also indicated that the growth in measurement error with increasing DBH may be exponential rather than linear, and that there are particularly fat tails of the error distribution suggesting a Laplace distribution better describes observed error rates than a Normal distribution (see Figure 3.6). However, I encountered computational difficulties employing a Laplace distribution within the observational model using this magnitude of measurement error and resorted to the best-fitting Normal distribution model for measurement error instead.

FIA measurements of radial increment recorded on a subset of trees at the second measurement visit (10 years after the first) were also used to build a second observation model. The five- and ten-year increment recorded at year 10 were used to condition the incremental growth predicted by the Wykoff process model, with measured radial increment assumed to be half of inside-bark diameter increment during that period. This is the DG observation model:

$$DG_t^* \sim N(DG_t, 0.03 \times DG_t) \quad (3.12)$$

Similar to the DBH observation model, the standard deviation of the DG observation model was assumed to increase linearly with the amount of growth recorded. Without a data-driven option to specify the measurement error for the DG observation model, I opted to utilize the FIA Measurement Quality Objective (MQO) which tolerates up to 0.05 inches of error per 1 inch of radial increment for 95% of measurements (USFS Forest Inventory and Analysis, 2023). The specification of the MQO and standard deviation for the DG observation model is perplexing because biometricians may expect measured diameter increment to be a more direct and reliable indicator of tree diameter growth over time than growth estimated from repeated DBH measurements with a diameter tape. Other studies which have utilized tree cores collected and measured precisely in laboratories with dendrochronological analyses have adopted observation models which treat increments or ring widths as much more precise observations than diameter tape measurements (Clark et al., 2007; Schliep et al., 2014; Heilman et al., 2022).

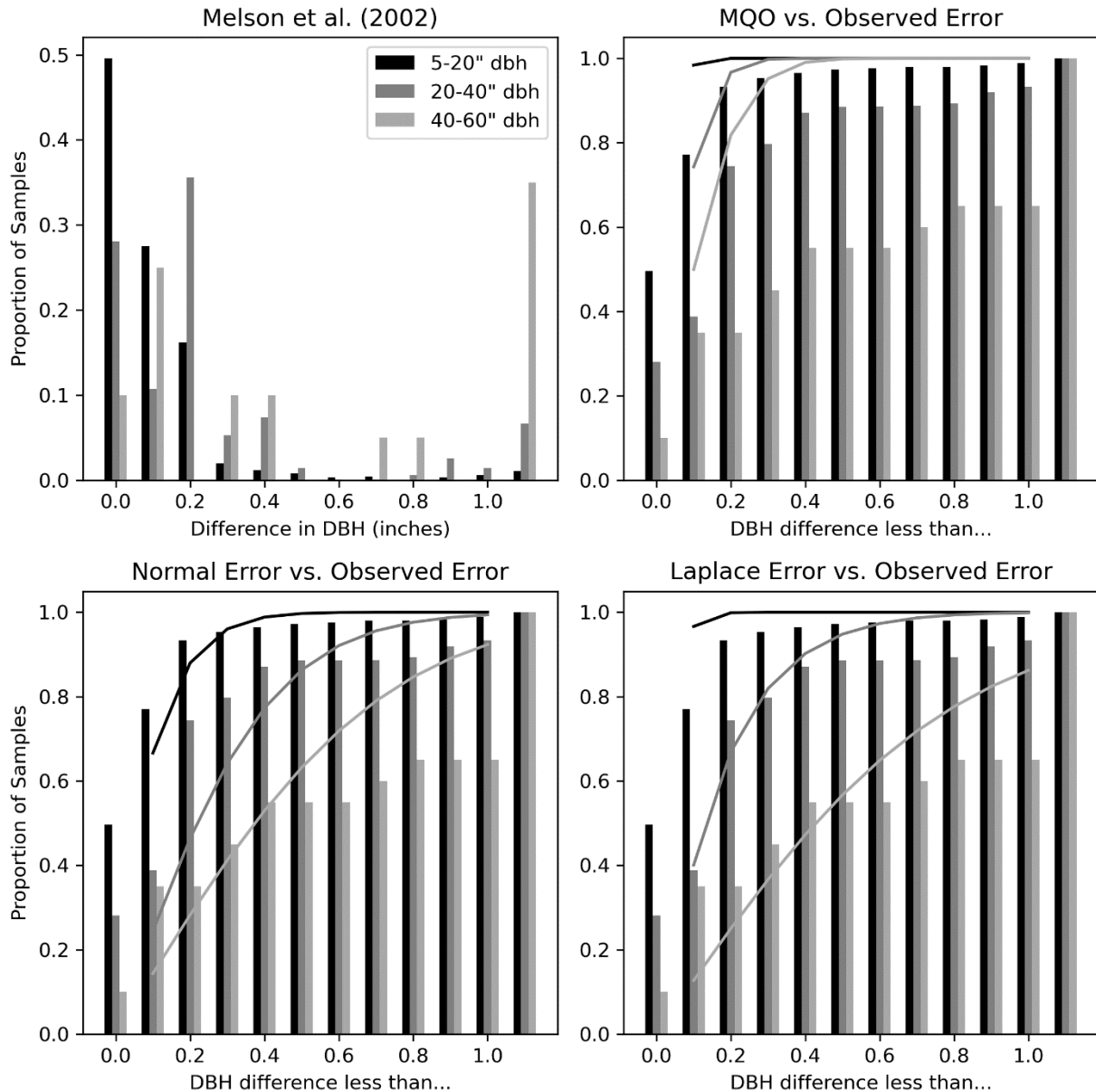


Figure 3.6. Adjustment to DBH measurement error

*Estimates of DBH measurement error originally reported in Melson et al. (2002) are shown in the top left graph. All other graphs translate the histogram of observed DBH differences in the top left graph into a cumulative distribution displayed as a bar plot. These cumulative distribution bar plots are then overplotted with lines displaying the proportion of trees that would be expected to have a measurement error of a particular magnitude for each diameter class. The top right graph displays the error rate expected if the FIA Measurement Quality Objective was being met. The lower left graph displays the best-fitting error rate using a Normal distribution with error linearly related to the measured DBH (scale = $0.011 * \text{DBH}$). The bottom right graph shows the best-fitting error rate using a Laplace distribution that has a non-linear relationship with DBH (scale = $0.00013 * \text{DBH}^{2.18}$)*

Moving from unpooled to hierarchical models

Diameter growth models were initially fit following the traditional (or “unpooled”) pattern used for developing FVS growth and mortality equations. That is, models for each species were fit with data from a single ecoregion or FVS variant. Although I implemented the fitting simultaneously across ecoregions for a single species, the unpooled model form treats the coefficients for the Wykoff process model in each ecoregion as independent from each other, and conditions the parameters in each regional model only on observations from within that ecoregion. These unpooled models were used to evaluate the potential for model simplification and consolidation and were intended to provide maximum flexibility for models to fit local data.

Following the selection of a preferred model form for each species and ecoregion, I developed a hierarchical or “partially pooled” model. To define a partially pooled model, each parameter in the Wykoff process model was defined to have each region-specific parameter drawn from a global parent distribution spanning all regions. The global distribution from which region-specific parameters were modeled with weakly informative Normal prior distributions (see Table 3.3).

Bayesian model fitting

State-Space models for simulating tree diameters over time were coded in Python using the Numpyro library (Phan et al., 2019). Models were fit using Markov Chain Monte Carlo (MCMC) using the No U-Turn Sampler (NUTS). Due to its development on top of the Jax framework (Bradbury et al., 2018), Numpyro allows execution of MCMC sampling using Graphical Processing Units (GPUs), enabling substantial speedups in computing compared to running MCMC on Central Processing Units (CPUs). MCMC chains were run for 1,000 warm-up samples (discarded) and the next 1,000 samples from the Markov Chain were retained for analyses. Traces of the MCMC samples were inspected visually, both in terms of individual traces and using pair plots for all combinations of model parameters to identify convergence or sampling issues. Several iterations of model refinement and adjustments to sampling algorithm arguments were completed to improve MCMC sampling performance. Convergence was evaluated using the Gelman-Rubin R-hat statistic and effective sample size to ensure satisfactory model fitting and sample size for representation of the posterior distribution of modeled parameters. As mentioned

Table 3.3. Parameterization of the State-Space Models

Predictor	Parameter	Pooled	Unpooled	Partially Pooled	
				μ	σ
intercept	b_0	$N(2.5, 0.25)$	$N(2.5, 0.25)$	$N(2.5, 0.1)$	$HN(1)$
$\ln(\text{DBH})$	b_1	$N(0, 1)$	$N(0.7, 0.25)$	$N(0.7, 0.1)$	$HN(0.1)$
DBH^{**2}	b_2	$N(0, 1)$	$N(-0.2, 0.25)$	$N(-0.2, 0.1)$	$HN(0.1)$
$\ln(\text{SI})$	b_3	$N(0, 1)$	$N(0.25, 0.25)$	$N(-0.25, 0.1)$	$HN(0.1)$
SL	b_4	$N(0, 1)$	$N(0.0, 0.25)$	$N(0, 0.1)$	$HN(0.1)$
SL^{**2}	b_5	$N(0, 1)$	$N(0.0, 0.25)$	$N(0, 0.1)$	$HN(0.1)$
$\text{SL} * \sin(\text{ASP})$	b_6	$N(0, 0.1)$	$N(0.0, 0.10)$	$N(0, 0.05)$	$HN(0.1)$
$\text{SL} * \cos(\text{ASP})$	b_7	$N(0, 0.1)$	$N(0.0, 0.10)$	$N(0, 0.05)$	$HN(0.1)$
EL	b_8	$N(0, 1)$	$N(-0.3, 0.25)$	$N(-0.3, 0.1)$	$HN(0.1)$
EL^{**2}	b_9	$N(0, 1)$	$N(0.15, 0.25)$	$N(0.15, 0.1)$	$HN(0.1)$
CR	b_{10}	$N(0, 1)$	$N(0.8, 0.25)$	$N(0.8, 0.1)$	$HN(0.1)$
CR^{**2}	b_{11}	$N(0, 1)$	$N(-0.4, 0.25)$	$N(-0.4, 0.1)$	$HN(0.1)$
TREE_COMP	b_{12}	$N(0, 1)$	$N(0, 0.25)$	$N(0, 0.25)$	$HN(0.1)$
STAND_COMP	b_{13}	$N(0, 1)$	$N(0, 0.25)$	$N(0, 0.25)$	$HN(0.1)$
FVS Location random effect	ϵ_{loc}		$N(0, 0.1)$		
FIA Plot random effect	ϵ_{plot}		$N(0, 0.1)$		
Process error	ϵ_{wyk}		$IG(10, 5)$		
Real DBH	DBH_t^*		$N(DBH_t, 0.01 * DBH_t)$		
Real DG	DG_t^*		$N(DG_t, 0.03 * DG_t)$		

Notes: Parameterizations shown above reflect probability distributions of Normal (N), Half Normal (HN), Inverse Gamma (IG). Distributions shown for b_1 through b_{13} values indicate the prior distributions used for those parameters. In the Pooled model a single parameter for each of b_1 through b_{13} was defined using these priors. In the Unpooled model, nine independent parameters (one for each FVS variant) were defined for each of b_1 through b_{13} using the same priors. In the Partially-Pooled model, nine covarying parameters (one for each FVS variant) were drawn for each of b_1 through b_{13} from a $Normal(\mu, \sigma)$.

briefly above, several computational and statistical adjustments were employed to reduce non-identifiability arising from random effects, prior distributions for parameters, and measurement errors.

All predictor variables were standardized to have zero mean and unit variance. A model was initially fit using a fully pooled approach for Douglas-fir, using observations across all ecoregions and without random effects. This fitting helped to identify plausible ranges for each fixed effect and process model error, which were translated into weakly-informative priors for each of these parameters that were later employed for the unpooled and partially pooled models

for all species. See Table 3.3 for the values used to define prior distributions for all parameters and model forms. In graphics displaying posterior distributions derived from MCMC sampling, prior distributions were also displayed to help illustrate the relative weakness or flatness of the chosen prior distributions in contrast to the estimated posterior distributions of these parameters.

Initial model checking indicated several correlated covariates, which can reduce the effectiveness of MCMC sampling (causing higher autocorrelation between subsequent samples in the Markov Chain and reducing sampling efficiency). Coefficients for the two size effects (b_1 and b_2) were highly correlated, as well as coefficients for tree- and stand-level competition variables (b_{12} and b_{13}). By specifying a dense mass matrix for these two combinations of variables as part of the Hamiltonian dynamics employed in the NUTS algorithm, sampling efficiency and convergence issues encountered when a diagonal mass matrix (default) was used were resolved. In the application of hierarchical models, a similar issue occurs due to the dependence between the parent and child parameters for each model coefficient occurs, and was addressed through the use of non-centered parameterizations (Papaspiliopoulos et al., 2003) for the coefficients for fixed effects in each ecoregion.

Performance assessment and model comparison

I calculated predictive performance for diameter growth predictions using mean absolute error (MAE), Root Mean Squared Error (RMSE), Bayesian R^2 (Gelman et al., 2019) and average bias metrics. I conducted model comparisons using the HDI-ROPE procedure introduced by Kruschke and colleagues (Kruschke, 2014; Kruschke and Liddell, 2018; Kruschke, 2018). In the context of model comparison, the HDI-ROPE decision rule involves defining a performance metric and a region of practical equivalence (ROPE) used to compare alternative models. I utilized MAE as the performance metric for the HDI-ROPE procedure and defined a ROPE for each tree species as 1% of the average 10-year DBH growth estimated from repeated DBH measurements in each region. I utilized the 90% highest density interval (HDI) of the posterior distribution (indicated in graphs as HPDI) of MAE to characterize the range of each model's predictive performance.

A priori, I identified a single model formulation as a preferred benchmark candidate: the SIMPLE-BALLNDBH-BA model. I chose this model as a preferred option because it can be calculated directly from inventory data without reliance on additional regression equations (as in the case of CCF) and does not involve arbitrary cutoffs (as in the case of RELHT). In this model,

the tree-level competition indicator ($BAL / \ln(DBH+1)$) is effectively an interaction term for tree- and stand-level competition. Finally, this model uses fewer parameters than the FULL-BALLNDBH-BA model.

The MAE for each model was calculated using 1,000 posterior samples of model coefficients and all FIA observations for that species. The difference in MAE for each of these 1,000 parameterizations was calculated, and the MAE of the candidate model was subtracted from the MAE of the benchmark model. I then calculated the 90% HDI for the difference in MAEs between models. As implemented here, the HDI-ROPE decision rule leads to one of the following outcomes for each pair of models compared (also illustrated in Figure 3.7):

- i. The candidate model is deemed inferior to the benchmark model if the upper bound of the 90% HDI falls below the lower bound of the ROPE. In other words, the candidate model has a larger MAE compared to the benchmark model that generally exceeds the ROPE.
- ii. The candidate model is deemed equivalent to the benchmark model if the HDI is fully enclosed within the ROPE. In other words, the candidate model rarely shows predictive performance that is better or worse than the benchmark model beyond the ROPE.
- iii. The candidate model is deemed superior to the benchmark model if the lower bound of the 90% HDI falls above the upper bound of the ROPE. In other words, the benchmark model has a larger MAE compared to the candidate model that generally exceeds the ROPE.
- iv. The comparison of models is deemed inconclusive if the 90% HDI encloses one or both of the ROPE boundaries. This is similar to the Frequentist concept of failing to reject a null hypothesis that two models have equivalent performance.

I conducted the HDI-ROPE procedure using the unpooled SIMPLE-BALLNDBH-BAL model as a benchmark. I used MAE of predictions of 10-year outside-bark diameter growth, and made the HDI-ROPE comparison within each FVS regional Variant. By performing model comparison within each FVS Variant, I attempted to determine whether there were regional differences in diameter growth that are better predicted using alternative competition variables or with the additional parameters in the FULL Wykoff model. If some regions were found where alternative models outperformed the benchmark model, it may justify the maintenance of a diversity of models across regions to best fit the available data.

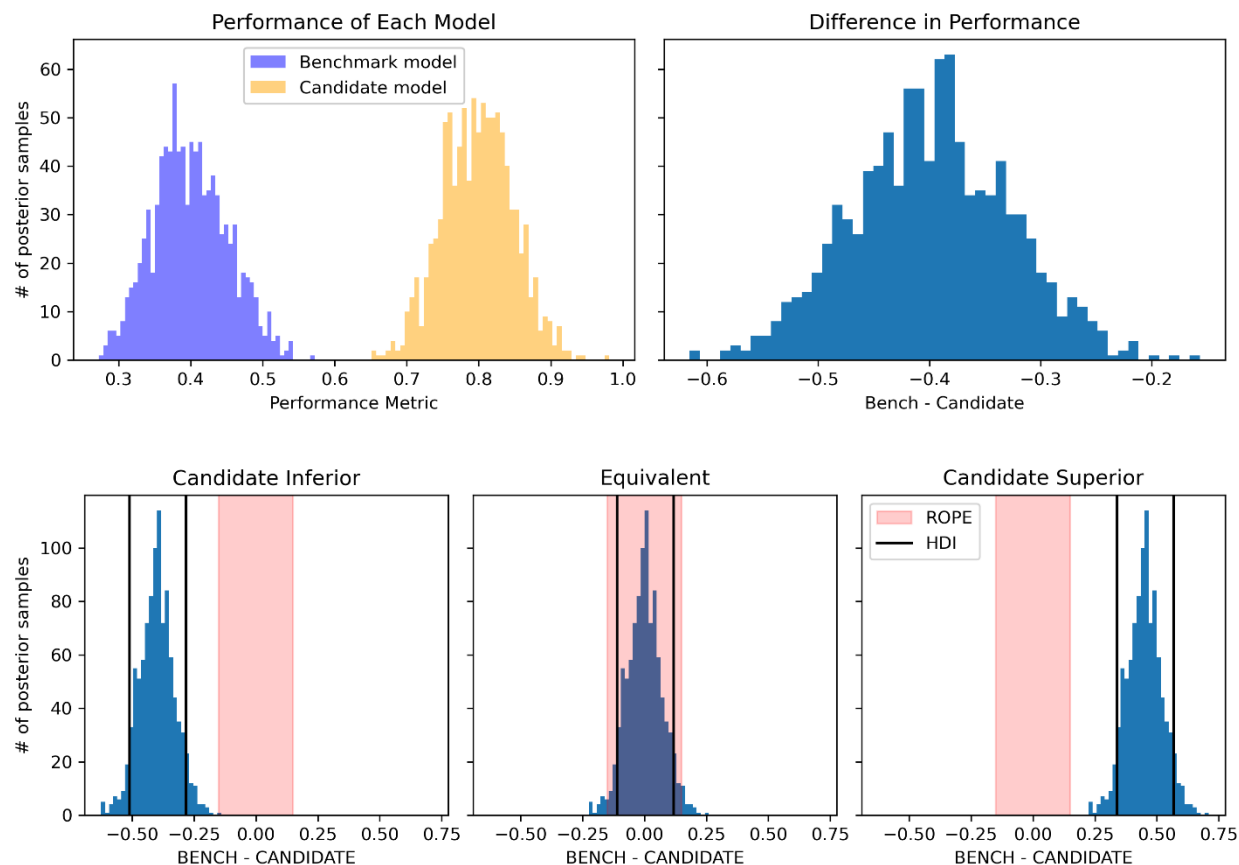


Figure 3.7. Illustration of HDI-ROPE decisions

The top row of graphs illustrates the calculation of a difference in performance metric used to compare models. The second row of graph illustrates how the HDI of that difference in performance metric between the two models is used with the ROPE to determine whether a candidate model is inferior, equivalent, or superior to the benchmark model. In instances where the HDI overlaps either of the lower or upper boundaries of the ROPE are deemed inconclusive (not shown in figure).

To evaluate the effects of employing a hierarchical model form, I also implemented the HDI-ROPE decision rule to compare the performance of a hierarchical implementation of the benchmark model against the unpooled version of the benchmark model.

Reproducible computing environment and workflows

All computing was performed on a Linux server running Ubuntu 20.04 equipped with 528GB of RAM and three NVIDIA GeForce RTX A400 Graphics Processing Units (GPUs), each with 16GB of GPU memory. The computing environment was installed using Miniconda including the Conda package manager (Conda contributors, 2022) and Python version 3.10. Model fitting

and analyses were implemented using a series of Jupyter Notebooks. All code and notebooks to implement the workflow are published on GitHub at https://github.com/d-diaz/refit_fvs under the permissive BSD 3-Clause free software license.

3.3 RESULTS

Unpooled Wykoff model comparisons

For each of the 15 species considered in this study, I compared 17 candidate models using the HDI-ROPE decision rule against the benchmark SIMPLE-BALLNDBH-BA model in each region where that species was observed across California, Oregon, and Washington. The results for the four focal FVS regional Variants of Westside Cascades (WC), Pacific Northwest Coast (PN), Inland California and Southern Cascades (CA), and Klamath Mountains (NC) are shown in Table 3.4. The results for the four focal FVS Variants of Westside Cascades (WC), Pacific Northwest Coast (PN), Inland California and Southern Cascades (CA), and Klamath Mountains (NC) are shown in Table 3.4. I did not observe a single instance in which a candidate model was judged superior to the benchmark model for any species in any ecoregion. For all 15 species and in all ecoregions, the majority of candidate models showed comparable performance in the prediction of diameter growth as the benchmark model, producing inconclusive results from the HDI-ROPE decision rule. For 8 of the 15 species considered, at least one candidate model was judged inferior to the benchmark model across the ecoregions considered, while the other 7 species showed all candidate models to either be inconclusive or equivalent to the benchmark model.

The ranking of models that would've occurred through point estimates of model performance was highly variable from species to species and among performance metrics. Several species showed clear breaks where the inclusion of specific features such as RELHT were consistently producing lower-performing models, but this was not true across all species. The comparison of all the unpooled models including the SIMPLE-BALLNDBH-BA benchmark and the 17 unpooled candidate models for Douglas-fir are displayed using multiple performance metrics in Figure 3.8. Similar visualizations for the 14 other tree species are displayed in Supplementary Materials.

Table 3.4. Summary of Unpooled Model Comparisons

Common Name	HDI-ROPE decisions within each FVS Regional Variant															
	Westside Cascades (WC)				Pacific Northwest Coast (PN)				Inland California and Southern Cascades (CA)				Klamath Mountains (NC)			
	?	<	=	>	?	<	=	>	?	<	=	>	?	<	=	>
Douglas-fir	6	7	4	0	8	7	2	0	10	7	0	0	10	7	0	0
Ponderosa pine	17	0	0	0	17	0	0	0	14	3	0	0	17	0	0	0
Western hemlock	11	4	2	0	9	6	2	0	17	0	0	0	17	0	0	0
White fir	17	0	0	0	17	0	0	0	17	0	0	0	17	0	0	0
Pacific silver fir	16	0	1	0	17	0	0	0	--	--	--	--	--	--	--	--
Mountain hemlock	16	1	0	0	17	0	0	0	17	0	0	0	17	0	0	0
Western redcedar	15	2	0	0	17	0	0	0	17	0	0	0	17	0	0	0
Canyon live oak	17	0	0	0	17	0	0	0	17	0	0	0	17	0	0	0
Tanoak	17	0	0	0	--	--	--	--	17	0	0	0	17	0	0	0
Red alder	17	0	0	0	12	5	0	0	17	0	0	0	17	0	0	0
California black oak	17	0	0	0	17	0	0	0	17	0	0	0	17	0	0	0
Pacific madrone	17	0	0	0	17	0	0	0	17	0	0	0	17	0	0	0
Coast redwood	--	--	--	--	--	--	--	--	--	--	--	--	14	3	0	0
Bigleaf maple	17	0	0	0	14	3	0	0	17	0	0	0	17	0	0	0
Noble fir	17	0	0	0	17	0	0	0	17	0	0	0	17	0	0	0

Notes: for each of the four focal regional variant of FVS (WC = Westside Cascades, PN = Pacific Northwest Coast, CA = Inland California and Southern Cascades, and NC = Klamath Mountains), 17 candidate models were compared using the HDI-ROPE decision rule against the SIMPLE-BALLNDBH-BAL benchmark model. The values in each cell correspond to the count of candidate models for a species in each ecoregion categorized as one of four model comparison outcomes: inconclusive (?); inferior than benchmark (<); equivalent to benchmark (=), and superior to benchmark (>). Regions where a species was never observed are indicated with "--".

Comparing the unpooled and partially pooled Wykoff benchmark models

For all 15 species and in all FVS regional variants across Oregon, Washington, and California, the performance of hierarchical or partially pooled versions of SIMPLE-BALLNDBH-BA models was comparable to the unpooled versions of the model, yielding inconclusive results from the HDI-ROPE decision rule.

Comparison with current FVS models

When compared against the default settings for the diameter growth models currently employed by FVS, refit models, both unpooled and hierarchical versions commonly outperformed FVS in terms of lower MAE and smaller bias (see Table 3.5 and Table 3.6). As a deterministic model, the performance estimates of FVS are points rather than distributions as is the case with

the Bayesian models. I applied the HDI-ROPE decision rule in this context by treating the FVS model in each ecoregion as the benchmark model for each species, and followed the HDI-ROPE determination procedure based on how the lower and upper bounds of the Bayesian model HDI for MAE fell in relation to the ROPE, which was again set at 1% of the average 10-year diameter growth observed for that species in that region. The comparisons between the hierarchical SIMPLE-BALLNDBH-BAL models and FVS are illustrated in Figure 3.9. FVS outperformed the benchmark hierarchical models in 5 out of 43 instances for: canyon live oak in the CA region; red alder in the NC region; California black oak in the CA and NC regions; and bigleaf maple in the PN region. The HDI-ROPE produced inconclusive comparisons in 7 out of 43 instances and showed the newly-fitted hierarchical models outperformed FVS in 31 out of 43 instances.

FVS generally showed more bias than the refit models for most species and regions. The average bias in FVS simulations was always positive, for every species and every region, with several regional models having an average bias of overpredicting 10-year diameter increment by more than 1 inch (see Table 3.5 and Table 3.6).

Posterior distributions of model parameters

All species showed similar patterns in terms of variable importance (see Table 3.7). Current tree size was the most influential driver of diameter growth, with both the log-transformed and quadratic terms having large coefficients. After tree size, live crown ratio was typically the next most influential factor on diameter growth, followed by the tree-level competition variable and Site Index. Slope and elevation were often among the smallest effects, with both features generally showing a negative effect on growth. Plot-level competition as indicated by basal area generally showed a small but positive effect on diameter growth. An illustration of the posterior distributions of model parameters for the partially pooled (hierarchical) SIMPLE-BALLDNBH-BA for Douglas-fir is shown in Figure 3.11. Similar plots for the other 14 species are included in Supplementary Materials.

When considering the magnitude of variation among ecoregion-level, location-level, and plot-level effects, hardwood species tended to show substantially higher degrees of ecoregion-level variation than softwoods (see Table 3.9). Mountain hemlock also showed substantial ecoregion-level variation among the intercept terms used in the hierarchical SIMPLE-

BALLNDBH-BA model. Most species showed comparable levels of variation among location- and plot-level parameters.

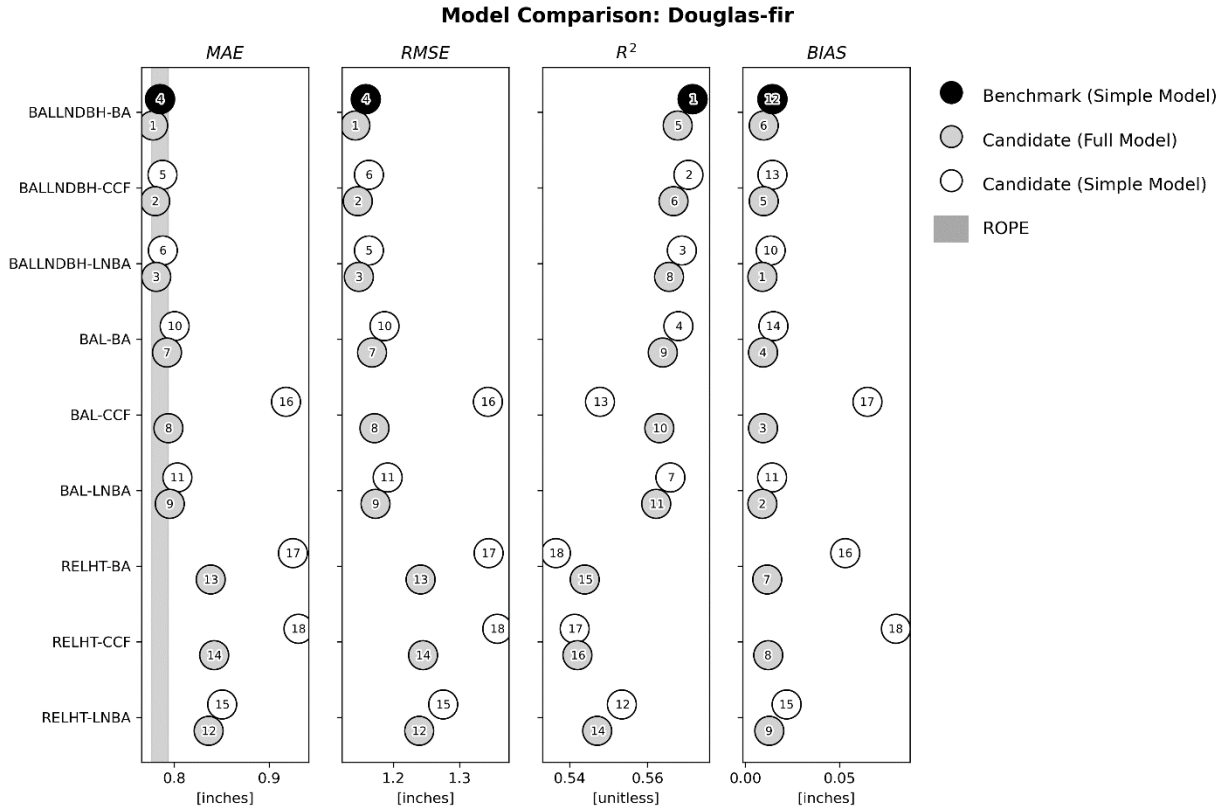


Figure 3.8. Unpooled model comparison with multiple performance metrics: Douglas-fir
These graphs illustrate the performance of the unpooled benchmark model (SIMPLE-BALLNDBH-BAL) as well as the 17 other unpooled candidate models using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Bayesian R^2 , and average bias. These performance metrics are calculated from all tree observations across the four focal FVS Variants. The combination of stand-level and tree-level competition variables used in each model are labeled at the left side of the graphs. The number inside each dot indicates the rank of the model based on the model's average for that performance metric across the four focal FVS Variants. The location of each dot along the x-axis indicates the mean value for the performance metric, and horizontal lines (if visible beyond the dot) indicate the standard deviation of the performance metric. The benchmark model is filled in black, while simple-form candidate models are hollow, and full-form candidate models are filled in gray. The vertical gray bar in the MAE graph is intended to illustrate the approximate size of the Region of Practical Equivalence employed in HDI-ROPE tests. However, it is worth noting the HDI-ROPE test was applied in each region independently (as opposed to using a single ROPE across all regions as shown above). The HDI-ROPE test was also applied using the 90% highest posterior density interval, as opposed to the mean $\pm$ the standard deviation.

Table 3.5. Performance metrics for FVS and refitted models: Hardwood species

Common Name	Model	<u>MAE</u>				<u>BIAS</u>				<u>R²</u>			
		WC	PN	CA	NC	WC	PN	CA	NC	WC	PN	CA	NC
Canyon live oak	FVS	--	--	0.330	0.346	--	--	-0.061	0.101	--	--	0.004	0.016
	Unpooled	--	--	0.369	0.363	--	--	0.003	0.021	--	--	0.353	0.401
	Hierarchical	--	--	0.368	0.363	--	--	-0.001	0.010	--	--	0.354	0.397
Tanoak	FVS	--	--	0.797	0.828	--	--	0.666	0.713	--	--	0.032	0.049
	Unpooled	--	--	0.484	0.626	--	--	0.056	0.048	--	--	0.417	0.459
	Hierarchical	--	--	0.472	0.626	--	--	0.031	0.043	--	--	0.413	0.457
Red alder	FVS	1.556	1.158	0.759	0.917	1.079	0.799	-0.576	-0.221	0.001	0.017	0.089	0.019
	Unpooled	0.913	1.050	0.936	1.383	0.111	0.076	0.162	0.179	0.563	0.498	0.443	0.457
	Hierarchical	0.912	1.052	0.876	1.328	0.110	0.075	0.105	0.091	0.562	0.497	0.428	0.441
California black oak	FVS	0.502	--	0.335	0.280	0.459	--	-0.077	0.052	0.004	--	0.004	0.062
	Unpooled	0.403	--	0.411	0.313	0.053	--	-0.011	0.012	0.403	--	0.289	0.377
	Hierarchical	0.383	--	0.413	0.313	-0.009	--	-0.016	0.001	0.342	--	0.286	0.368
Pacific madrone	FVS	0.553	0.815	0.594	0.772	0.085	0.707	0.172	0.658	0.010	0.036	0.005	0.052
	Unpooled	0.431	0.515	0.594	0.544	0.029	0.095	-0.004	0.034	0.476	0.476	0.220	0.460
	Hierarchical	0.425	0.486	0.593	0.551	0.006	0.030	-0.015	0.025	0.469	0.424	0.214	0.460
Bigleaf maple	FVS	1.087	1.205	0.355	0.868	0.587	0.559	0.028	0.648	0.035	0.027	0.003	0.000
	Unpooled	0.920	1.377	0.426	0.853	0.069	0.123	0.067	0.207	0.487	0.455	0.399	0.437
	Hierarchical	0.912	1.370	0.402	0.810	0.042	0.102	0.016	0.096	0.476	0.453	0.381	0.423

Notes: Values in the table shown for “Unpooled” and “Hierarchical” indicate the average MAE, BIAS, and R² for the unpooled and hierarchical versions of the SIMPLE-BALLNDBH-BA model, respectively. Values are omitted (“--”) for regions and species where there were fewer than 10 observations.

Table 3.6. Performance metrics for FVS and refitted models: Softwood species

Common Name	Model	<u>MAE</u>				<u>BIAS</u>				<u>R2</u>			
		WC	PN	CA	NC	WC	PN	CA	NC	WC	PN	CA	NC
Douglas-fir	FVS	0.741	0.989	0.742	0.905	0.011	0.028	-0.001	0.018	0.606	0.534	0.521	0.541
	Unpooled	0.742	0.990	0.742	0.904	0.010	0.030	-0.002	0.015	0.606	0.534	0.521	0.540
	Hierarchical	2.143	--	1.126	1.233	1.771	--	0.724	0.831	0.291	--	0.235	0.125
Ponderosa pine	FVS	0.925	--	0.749	0.674	-0.049	--	-0.070	0.050	0.592	--	0.545	0.545
	Unpooled	0.929	--	0.745	0.680	-0.060	--	-0.079	0.059	0.586	--	0.540	0.535
	Hierarchical	0.645	1.305	0.973	1.671	0.107	1.056	0.777	1.483	0.148	0.277	0.309	0.119
Western hemlock	FVS	0.644	0.906	0.522	0.777	0.023	0.040	0.008	0.041	0.567	0.535	0.561	0.487
	Unpooled	0.644	0.906	0.512	0.755	0.022	0.040	-0.042	0.039	0.566	0.535	0.555	0.487
	Hierarchical	0.820	--	0.948	1.057	0.488	--	0.691	0.784	0.168	--	0.180	0.266
White fir	FVS	0.656	--	0.690	0.560	0.003	--	-0.003	0.017	0.558	--	0.507	0.502
	Unpooled	0.653	--	0.691	0.560	0.003	--	-0.002	0.034	0.559	--	0.506	0.505
	Hierarchical	0.754	0.913	--	--	0.333	0.580	--	--	0.176	0.124	--	--
Pacific silver fir	FVS	0.566	0.727	--	--	0.015	0.008	--	--	0.591	0.482	--	--
	Unpooled	0.566	0.725	--	--	0.015	0.006	--	--	0.590	0.482	--	--
	Hierarchical	0.741	0.989	0.742	0.905	0.011	0.028	-0.001	0.018	0.606	0.534	0.521	0.541
Mountain hemlock	FVS	0.345	0.497	1.086	--	0.101	0.278	0.738	--	0.114	0.009	0.115	--
	Unpooled	0.362	0.353	0.325	--	-0.000	0.047	0.015	--	0.459	0.210	0.508	--
	Hierarchical	0.361	0.365	0.321	--	-0.004	0.024	-0.015	--	0.453	0.196	0.479	--
Western redcedar	FVS	0.764	1.305	--	--	0.097	0.860	--	--	0.076	0.061	--	--
	Unpooled	0.802	1.203	--	--	0.039	0.060	--	--	0.505	0.475	--	--
	Hierarchical	0.802	1.199	--	--	0.036	0.059	--	--	0.503	0.475	--	--
Coast redwood	FVS	--	--	--	1.515	--	--	--	0.914	--	--	--	0.055
	Unpooled	--	--	--	1.349	--	--	--	0.061	--	--	--	0.489
	Hierarchical	--	--	--	1.349	--	--	--	0.063	--	--	--	0.489
Noble fir	FVS	1.102	2.210	--	--	0.330	2.495	--	--	0.149	0.152	--	--
	Unpooled	0.813	1.104	--	--	0.031	0.192	--	--	0.627	0.658	--	--
	Hierarchical	0.814	1.014	--	--	0.033	0.090	--	--	0.627	0.647	--	--

Notes: See explanatory notes in Table 3.5.

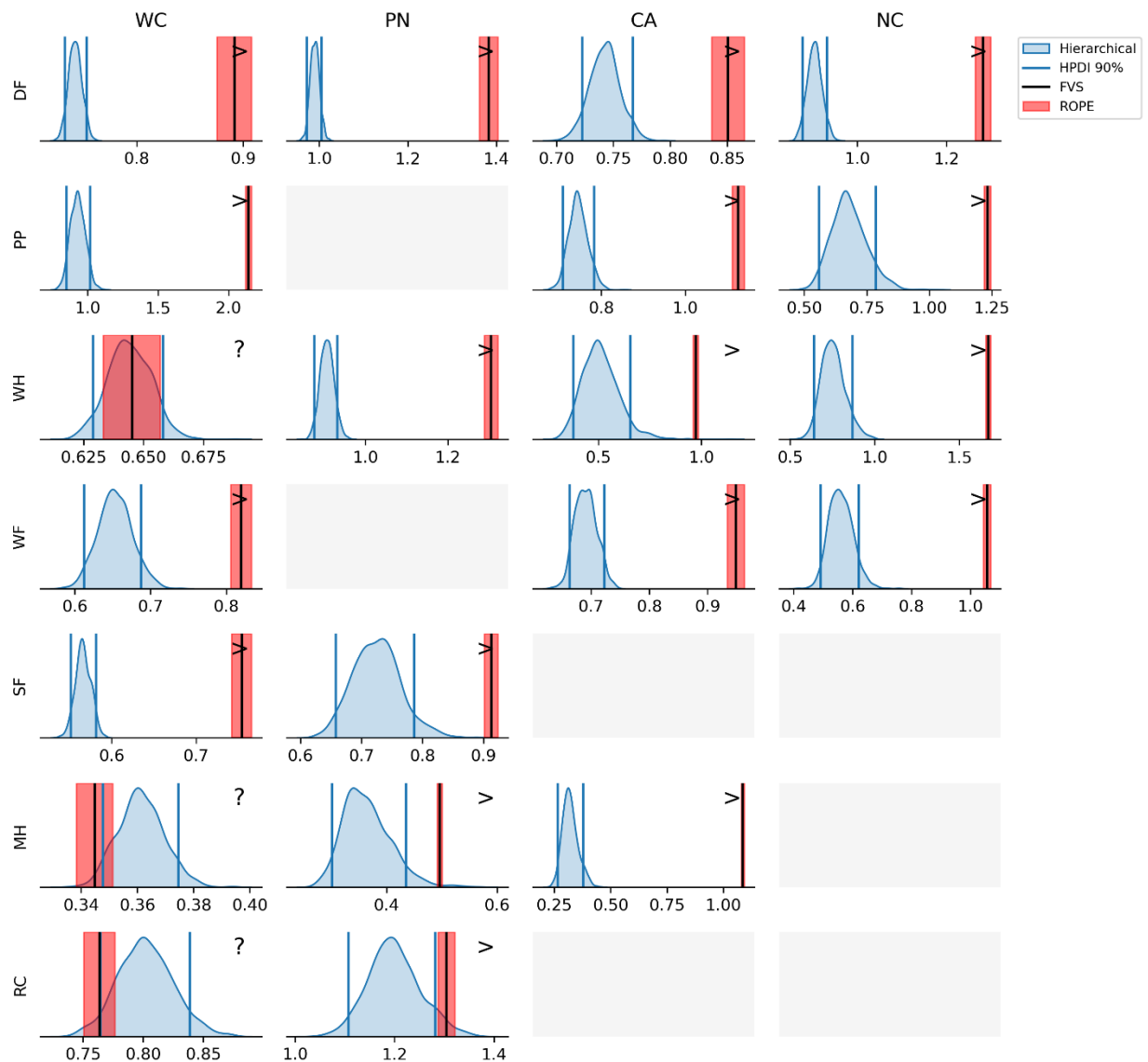


Figure 3.9. HDI-ROPE comparisons between FVS default models and hierarchical benchmark models, Part 1 of 2.

Each row of graphs shows a kernel density estimate of the distribution of the Mean Absolute Error for 10-year diameter growth predictions for single species across four regions (WC = Westside Cascades, PN = Pacific Northwest Coast, CA = Inland California and Southern Cascades, NC = Klamath Mountains). The MAE of FVS and ROPE, are shown as vertical line and red band, respectively. The outcome of the HDI-ROPE decision rule for each species in each region is indicated in the upper-right corner of each graph (>: superior to FVS, ?: inconclusive comparison with FVS, <: inferior to FVS). Species codes are: DF = Douglas-fir; PP = Ponderosa pine; WH = Western hemlock; WF = White fir; SF = Pacific silver fir; MH = Mountain hemlock; RC = Western redcedar.

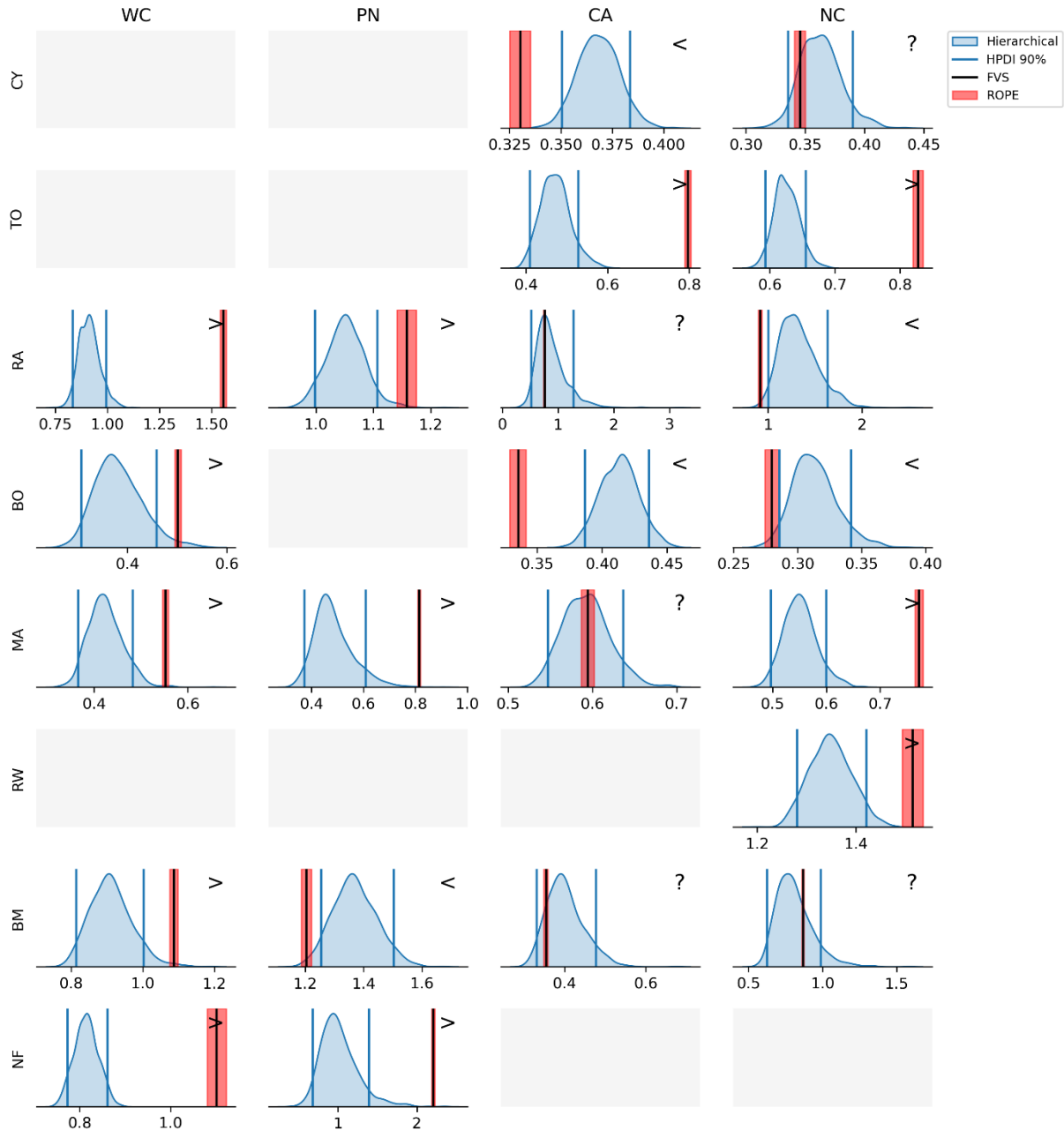


Figure 3.10. HDI-ROPE comparisons between FVS default models and hierarchical benchmark models, Part 2 of 2.

Each row of graphs shows a kernel density estimate of the distribution of the Mean Absolute Error for 10-year diameter growth predictions for single species across four regions (WC = Westside Cascades, PN = Pacific Northwest Coast, CA = Inland California and Southern Cascades, NC = Klamath Mountains). The MAE of FVS and ROPE, are shown as vertical line and red band, respectively. The outcome of the HDI-ROPE decision rule for each species in each region is indicated in the upper-right corner of each graph (>: superior to FVS, ?: inconclusive comparison with FVS, <: inferior to FVS). Species codes are: CY = Canyon live oak; TO = Tanoak; RA = Red alder; BO = California black oak; MA = Pacific madrone; RW = Coast redwood; BM = Bigleaf maple; NF = Noble fir.

Table 3.7. Distributions of standardized model coefficients for hierarchical SIMPLE-BALLNDBH-BA models

SPP.	VAR.	b_0	b_1	b_2	b_3	b_4	b_8	b_{10}	b_{12}	b_{13}	ε_{wyk}
		<i>intercept</i>	<i>ln(DBH)</i>	<i>DBH**2</i>	<i>ln(SI)</i>	<i>SL</i>	<i>EL</i>	<i>CR</i>	<i>BAL/ ln(DBH+1)</i>	<i>BA</i>	<i>process error</i>
DF	WC	2.681 ± 0.02	0.669 ± 0.01	-0.262 ± 0.01	0.269 ± 0.01	-0.050 ± 0.01	-0.069 ± 0.01	0.453 ± 0.01	-0.320 ± 0.01	0.117 ± 0.01	0.576 ± 0.00
	PN	2.598 ± 0.03	0.697 ± 0.01	-0.160 ± 0.01	0.187 ± 0.01	-0.041 ± 0.01	-0.177 ± 0.01	0.423 ± 0.01	-0.392 ± 0.01	0.104 ± 0.01	0.614 ± 0.01
	CA	2.579 ± 0.02	0.659 ± 0.02	-0.391 ± 0.03	0.188 ± 0.01	0.007 ± 0.01	-0.078 ± 0.02	0.392 ± 0.01	-0.351 ± 0.01	0.176 ± 0.02	0.658 ± 0.01
	NC	2.553 ± 0.03	0.681 ± 0.02	-0.299 ± 0.02	0.245 ± 0.01	-0.068 ± 0.01	-0.209 ± 0.02	0.428 ± 0.01	-0.367 ± 0.01	0.177 ± 0.02	0.691 ± 0.01
PP	WC	2.312 ± 0.06	0.659 ± 0.03	-0.357 ± 0.03	0.128 ± 0.03	-0.124 ± 0.04	-0.223 ± 0.04	0.384 ± 0.03	-0.268 ± 0.02	0.058 ± 0.02	0.624 ± 0.03
	PN	2.319 ± 0.13	0.646 ± 0.04	-0.348 ± 0.06	0.177 ± 0.06	-0.067 ± 0.08	-0.177 ± 0.07	0.396 ± 0.06	-0.273 ± 0.02	0.061 ± 0.05	0.514 ± 0.17
	CA	2.255 ± 0.04	0.665 ± 0.03	-0.317 ± 0.03	0.170 ± 0.02	-0.076 ± 0.02	-0.051 ± 0.02	0.439 ± 0.02	-0.277 ± 0.01	0.082 ± 0.02	0.668 ± 0.02
	NC	2.229 ± 0.08	0.648 ± 0.03	-0.370 ± 0.05	0.161 ± 0.03	0.009 ± 0.05	-0.066 ± 0.05	0.374 ± 0.04	-0.267 ± 0.02	0.064 ± 0.03	0.689 ± 0.06
WH	WC	2.298 ± 0.02	0.581 ± 0.02	-0.275 ± 0.02	0.135 ± 0.01	0.003 ± 0.01	-0.112 ± 0.01	0.491 ± 0.01	-0.421 ± 0.02	0.174 ± 0.02	0.679 ± 0.01
	PN	2.286 ± 0.03	0.620 ± 0.02	-0.266 ± 0.02	0.083 ± 0.02	0.011 ± 0.01	-0.227 ± 0.02	0.418 ± 0.01	-0.502 ± 0.02	0.163 ± 0.02	0.722 ± 0.01
	CA	2.411 ± 0.12	0.608 ± 0.06	-0.277 ± 0.05	0.086 ± 0.06	0.003 ± 0.04	-0.210 ± 0.15	0.446 ± 0.06	-0.402 ± 0.07	0.179 ± 0.04	0.488 ± 0.07
	NC	2.263 ± 0.08	0.609 ± 0.05	-0.281 ± 0.05	0.092 ± 0.04	-0.005 ± 0.03	-0.164 ± 0.08	0.446 ± 0.05	-0.330 ± 0.05	0.192 ± 0.04	0.721 ± 0.05
WF	WC	2.388 ± 0.05	0.803 ± 0.03	-0.357 ± 0.03	0.118 ± 0.02	-0.054 ± 0.01	-0.074 ± 0.03	0.437 ± 0.02	-0.332 ± 0.03	0.143 ± 0.03	0.597 ± 0.02
	PN	2.467 ± 0.12	0.791 ± 0.06	-0.356 ± 0.07	0.130 ± 0.04	-0.057 ± 0.03	-0.137 ± 0.05	0.380 ± 0.09	-0.311 ± 0.08	0.128 ± 0.06	0.506 ± 0.17
	CA	2.385 ± 0.03	0.788 ± 0.03	-0.336 ± 0.03	0.149 ± 0.02	-0.048 ± 0.01	-0.143 ± 0.02	0.333 ± 0.02	-0.353 ± 0.03	0.122 ± 0.02	0.603 ± 0.01
	NC	2.371 ± 0.06	0.770 ± 0.05	-0.370 ± 0.05	0.123 ± 0.03	-0.069 ± 0.03	-0.071 ± 0.04	0.317 ± 0.04	-0.276 ± 0.04	0.116 ± 0.04	0.549 ± 0.04
SF	WC	1.970 ± 0.03	0.713 ± 0.02	-0.343 ± 0.03	0.118 ± 0.01	0.040 ± 0.01	-0.121 ± 0.01	0.567 ± 0.01	-0.267 ± 0.02	0.117 ± 0.02	0.603 ± 0.01
	PN	1.878 ± 0.10	0.718 ± 0.04	-0.226 ± 0.04	0.078 ± 0.04	-0.085 ± 0.04	-0.119 ± 0.03	0.534 ± 0.04	-0.361 ± 0.05	0.114 ± 0.04	0.793 ± 0.03
	CA	--	--	--	--	--	--	--	--	--	--
	NC	--	--	--	--	--	--	--	--	--	--
MH	WC	1.682 ± 0.04	0.655 ± 0.04	-0.197 ± 0.03	0.161 ± 0.02	0.024 ± 0.02	0.011 ± 0.03	0.422 ± 0.02	-0.299 ± 0.03	0.094 ± 0.02	0.646 ± 0.02
	PN	1.305 ± 0.22	0.684 ± 0.06	-0.147 ± 0.07	0.120 ± 0.06	0.002 ± 0.05	-0.009 ± 0.12	0.365 ± 0.07	-0.269 ± 0.06	0.151 ± 0.08	0.752 ± 0.10
	CA	1.666 ± 0.15	0.695 ± 0.06	-0.148 ± 0.08	0.170 ± 0.05	0.020 ± 0.04	-0.159 ± 0.18	0.412 ± 0.04	-0.254 ± 0.05	0.120 ± 0.05	0.614 ± 0.06
	NC	1.472 ± 0.30	0.697 ± 0.07	-0.189 ± 0.11	0.148 ± 0.09	0.014 ± 0.04	-0.060 ± 0.16	0.419 ± 0.06	-0.253 ± 0.07	0.114 ± 0.06	0.693 ± 0.19

Table 3.7. Distributions of standardized model coefficients for hierarchical SIMPLE-BALLNDBH-BA models

		b₀	b₁	b₂	b₃	b₄	b₈	b₁₀	b₁₂	b₁₃	ε_{wyk}
SPP.	VAR.	<i>intercept</i>	<i>ln(DBH)</i>	<i>DBH**2</i>	<i>ln(SI)</i>	<i>SL</i>	<i>EL</i>	<i>CR</i>	<i>BAL/ ln(DBH+1)</i>	<i>BA</i>	<i>process error</i>
RC	WC	2.516 ± 0.03	0.814 ± 0.04	-0.186 ± 0.03	0.073 ± 0.02	-0.019 ± 0.02	-0.156 ± 0.03	0.502 ± 0.02	-0.368 ± 0.03	0.131 ± 0.03	0.748 ± 0.02
	PN	2.049 ± 0.08	0.881 ± 0.05	-0.156 ± 0.04	0.036 ± 0.03	0.089 ± 0.03	-0.325 ± 0.05	0.487 ± 0.03	-0.386 ± 0.04	0.112 ± 0.04	0.883 ± 0.03
	CA	2.395 ± 0.19	0.803 ± 0.08	-0.193 ± 0.08	-0.089 ± 0.11	-0.032 ± 0.10	-0.272 ± 0.15	0.499 ± 0.05	-0.398 ± 0.10	0.079 ± 0.09	0.526 ± 0.15
	NC	2.569 ± 0.23	0.808 ± 0.09	-0.191 ± 0.08	0.098 ± 0.14	-0.019 ± 0.10	-0.303 ± 0.13	0.520 ± 0.07	-0.334 ± 0.13	0.108 ± 0.09	0.485 ± 0.15
CY	WC	1.343 ± 0.25	0.436 ± 0.12	-0.100 ± 0.08	0.134 ± 0.13	-0.064 ± 0.07	-0.200 ± 0.17	0.650 ± 0.20	-0.306 ± 0.06	0.133 ± 0.09	0.458 ± 0.12
	PN	0.519 ± 0.58	0.452 ± 0.13	-0.100 ± 0.08	0.117 ± 0.13	-0.069 ± 0.08	-0.148 ± 0.16	0.715 ± 0.19	-0.290 ± 0.07	0.162 ± 0.10	0.524 ± 0.17
	CA	0.863 ± 0.04	0.388 ± 0.04	-0.070 ± 0.04	0.082 ± 0.03	-0.039 ± 0.02	-0.061 ± 0.03	0.248 ± 0.02	-0.290 ± 0.04	0.153 ± 0.04	0.872 ± 0.02
	NC	0.764 ± 0.07	0.368 ± 0.05	-0.100 ± 0.05	0.198 ± 0.04	-0.064 ± 0.03	-0.057 ± 0.04	0.336 ± 0.04	-0.290 ± 0.04	0.195 ± 0.05	0.923 ± 0.04
TO	WC	1.868 ± 0.44	0.522 ± 0.10	-0.168 ± 0.11	0.252 ± 0.20	0.023 ± 0.08	-0.309 ± 0.21	0.396 ± 0.22	-0.527 ± 0.12	0.281 ± 0.11	1.038 ± 0.30
	PN	--	--	--	--	--	--	--	--	--	--
	CA	1.536 ± 0.11	0.504 ± 0.07	-0.133 ± 0.07	-0.095 ± 0.07	0.030 ± 0.04	-0.011 ± 0.09	0.165 ± 0.06	-0.463 ± 0.08	0.325 ± 0.07	0.817 ± 0.06
	NC	1.259 ± 0.04	0.488 ± 0.05	-0.147 ± 0.04	0.106 ± 0.02	0.013 ± 0.02	-0.042 ± 0.02	0.382 ± 0.02	-0.579 ± 0.05	0.286 ± 0.05	1.018 ± 0.02
RA	WC	2.213 ± 0.06	0.263 ± 0.08	-0.219 ± 0.10	0.149 ± 0.03	-0.025 ± 0.03	-0.057 ± 0.04	0.586 ± 0.03	-0.221 ± 0.05	0.095 ± 0.05	0.854 ± 0.03
	PN	2.424 ± 0.05	0.036 ± 0.05	-0.095 ± 0.04	0.043 ± 0.02	-0.015 ± 0.02	0.037 ± 0.03	0.404 ± 0.03	-0.425 ± 0.03	0.156 ± 0.04	0.862 ± 0.02
	CA	2.395 ± 0.16	0.195 ± 0.17	-0.205 ± 0.12	0.058 ± 0.13	-0.019 ± 0.07	-0.162 ± 0.12	0.628 ± 0.14	-0.255 ± 0.13	0.073 ± 0.10	0.725 ± 0.15
	NC	2.439 ± 0.12	0.139 ± 0.12	-0.144 ± 0.10	0.052 ± 0.10	0.008 ± 0.06	-0.144 ± 0.12	0.625 ± 0.10	-0.131 ± 0.10	0.120 ± 0.08	0.864 ± 0.09
BO	WC	1.471 ± 0.13	0.646 ± 0.08	-0.047 ± 0.08	-0.004 ± 0.05	0.033 ± 0.06	-0.146 ± 0.10	0.262 ± 0.09	-0.141 ± 0.05	0.010 ± 0.07	0.762 ± 0.08
	PN	1.095 ± 0.48	0.645 ± 0.09	-0.090 ± 0.08	0.017 ± 0.07	-0.003 ± 0.09	-0.158 ± 0.11	0.493 ± 0.25	-0.144 ± 0.07	0.014 ± 0.10	0.511 ± 0.18
	CA	1.335 ± 0.04	0.586 ± 0.06	-0.054 ± 0.06	0.017 ± 0.02	0.057 ± 0.03	-0.125 ± 0.04	0.221 ± 0.03	-0.155 ± 0.05	-0.030 ± 0.05	0.777 ± 0.03
	NC	1.289 ± 0.08	0.713 ± 0.07	-0.104 ± 0.06	0.001 ± 0.03	-0.003 ± 0.04	-0.099 ± 0.05	0.236 ± 0.05	-0.140 ± 0.05	0.096 ± 0.06	0.634 ± 0.04
MA	WC	1.384 ± 0.09	0.438 ± 0.07	-0.186 ± 0.10	0.066 ± 0.06	-0.076 ± 0.04	0.071 ± 0.09	0.443 ± 0.04	-0.371 ± 0.06	0.213 ± 0.06	0.832 ± 0.06
	PN	1.311 ± 0.28	0.482 ± 0.09	-0.179 ± 0.11	0.079 ± 0.11	-0.061 ± 0.06	-0.015 ± 0.13	0.477 ± 0.08	-0.369 ± 0.07	0.233 ± 0.08	1.012 ± 0.11
	CA	1.441 ± 0.07	0.472 ± 0.05	-0.178 ± 0.08	0.244 ± 0.05	-0.093 ± 0.03	0.005 ± 0.04	0.436 ± 0.03	-0.354 ± 0.04	0.231 ± 0.04	0.938 ± 0.03
	NC	1.229 ± 0.06	0.488 ± 0.06	-0.104 ± 0.05	0.069 ± 0.03	-0.085 ± 0.03	-0.111 ± 0.04	0.416 ± 0.04	-0.362 ± 0.05	0.200 ± 0.05	0.922 ± 0.04

Table 3.7. Distributions of standardized model coefficients for hierarchical SIMPLE-BALLNDBH-BA models

SPP.	VAR.	b_0 <i>intercept</i>	b_1 <i>ln(DBH)</i>	b_2 <i>DBH**2</i>	b_3 <i>ln(SI)</i>	b_4 <i>SL</i>	b_8 <i>EL</i>	b_{10} <i>CR</i>	b_{12} <i>BAL/ ln(DBH+1)</i>	b_{13} <i>BA</i>	ϵ_{wyk} <i>process error</i>
RW	WC	--	--	--	--	--	--	--	--	--	--
	PN	--	--	--	--	--	--	--	--	--	--
	CA	--	--	--	--	--	--	--	--	--	--
	NC	2.840 ± 0.03	0.934 ± 0.04	-0.382 ± 0.07	0.180 ± 0.02	0.011 ± 0.02	0.059 ± 0.02	0.385 ± 0.02	-0.468 ± 0.03	0.133 ± 0.04	0.840 ± 0.02
BM	WC	1.865 ± 0.07	0.382 ± 0.07	-0.199 ± 0.07	0.062 ± 0.07	-0.030 ± 0.04	-0.248 ± 0.05	0.385 ± 0.06	-0.487 ± 0.07	0.188 ± 0.08	1.088 ± 0.05
	PN	1.954 ± 0.09	0.411 ± 0.06	-0.171 ± 0.05	0.091 ± 0.07	-0.019 ± 0.04	-0.307 ± 0.10	0.412 ± 0.04	-0.544 ± 0.06	0.228 ± 0.06	1.052 ± 0.04
	CA	1.193 ± 0.15	0.409 ± 0.09	-0.204 ± 0.09	-0.027 ± 0.05	-0.025 ± 0.07	-0.117 ± 0.08	0.201 ± 0.12	-0.441 ± 0.10	0.204 ± 0.10	0.893 ± 0.08
	NC	1.193 ± 0.13	0.476 ± 0.09	-0.182 ± 0.08	0.061 ± 0.06	0.022 ± 0.08	-0.199 ± 0.09	0.394 ± 0.12	-0.425 ± 0.11	0.307 ± 0.11	1.207 ± 0.10
NF	WC	2.869 ± 0.04	0.919 ± 0.04	-0.281 ± 0.04	0.069 ± 0.02	0.001 ± 0.02	-0.103 ± 0.02	0.581 ± 0.02	-0.198 ± 0.02	0.084 ± 0.03	0.494 ± 0.01
	PN	2.675 ± 0.15	0.885 ± 0.09	-0.264 ± 0.09	0.218 ± 0.11	0.031 ± 0.09	-0.142 ± 0.05	0.652 ± 0.09	-0.019 ± 0.10	0.154 ± 0.08	0.554 ± 0.11
	CA	2.611 ± 0.27	0.861 ± 0.11	-0.266 ± 0.09	0.171 ± 0.12	0.020 ± 0.08	-0.145 ± 0.08	0.655 ± 0.11	-0.070 ± 0.14	0.116 ± 0.09	0.534 ± 0.19
	NC	2.739 ± 0.26	0.864 ± 0.11	-0.271 ± 0.09	0.169 ± 0.12	0.008 ± 0.09	-0.130 ± 0.08	0.650 ± 0.10	-0.053 ± 0.14	0.116 ± 0.11	0.575 ± 0.20

Notes: Each cell in this table displays the mean and standard deviation of the posterior distribution for each model parameter. Because the predictor variables were all standardized to have a zero mean and unit variance prior to model fitting, the magnitude of each parameter is a helpful indicator of the relative sensitivity of diameter growth to that variable. This table displays summary statistics of parameter values visualized in Figure 3.11 and similar figures for 14 other tree species included in Supplementary Materials. Regression coefficients that can be used with non-standardized covariate values are shown in Table 3.34. Cells with ‘—’ indicate regions where there were no samples of that species.

Table 3.8. Mean values of unstandardized model coefficients for hierarchical SIMPLE-BALLNDBH-BA models

		b₀	b₁	b₂	b₃	b₄	b₈	b₁₀	b₁₂	b₁₃
SPP.	VAR.	<i>intercept</i>	<i>ln(DBH)</i>	<i>DBH**2</i>	<i>ln(MAI)</i>	<i>SL</i>	<i>EL</i>	<i>CR</i>	<i>BAL/ ln(DBH+1)</i>	<i>BA</i>
DF	WC	-2.715	0.977	-3.606e-4	0.478	-0.206	-4.771e-3	2.222	-0.035	3.595e-3
	PN	-1.947	1.018	-2.192e-4	0.333	-0.169	-1.230e-2	2.074	-0.043	3.198e-3
	CA	-2.016	0.963	-5.375e-4	0.335	0.028	-5.432e-3	1.923	-0.038	5.402e-3
	NC	-2.357	0.995	-4.111e-4	0.436	-0.278	-1.451e-2	2.096	-0.040	5.421e-3
PP	WC	-1.263	1.106	-9.313e-4	0.277	-0.715	-2.246e-2	2.174	-0.044	3.165e-3
	PN	-1.956	1.084	-9.068e-4	0.383	-0.385	-1.783e-2	2.241	-0.045	3.313e-3
	CA	-2.810	1.116	-8.270e-4	0.367	-0.440	-5.097e-3	2.486	-0.045	4.460e-3
	NC	-2.426	1.088	-9.641e-4	0.348	0.053	-6.643e-3	2.114	-0.043	3.492e-3
WH	WC	-2.029	0.954	-5.584e-4	0.259	0.011	-9.203e-3	2.447	-0.029	5.257e-3
	PN	-1.187	1.018	-5.410e-4	0.159	0.044	-1.855e-2	2.082	-0.034	4.916e-3
	CA	-1.275	0.998	-5.623e-4	0.164	0.013	-1.723e-2	2.221	-0.027	5.387e-3
	NC	-1.682	1.000	-5.700e-4	0.177	-0.020	-1.342e-2	2.221	-0.022	5.786e-3
WF	WC	-2.239	1.261	-7.888e-4	0.206	-0.307	-6.621e-3	2.215	-0.027	4.869e-3
	PN	-1.712	1.243	-7.847e-4	0.227	-0.324	-1.230e-2	1.923	-0.026	4.350e-3
	CA	-1.750	1.239	-7.413e-4	0.260	-0.268	-1.279e-2	1.685	-0.029	4.162e-3
	NC	-1.817	1.210	-8.155e-4	0.216	-0.390	-6.355e-3	1.607	-0.023	3.966e-3
SF	WC	-2.445	1.221	-8.696e-4	0.199	0.171	-1.226e-2	2.574	-0.019	3.598e-3
	PN	-1.944	1.230	-5.736e-4	0.131	-0.361	-1.209e-2	2.422	-0.025	3.501e-3
	CA	--	--	--	--	--	--	--	--	--
	NC	--	--	--	--	--	--	--	--	--
MH	WC	-2.919	1.067	-4.681e-4	0.224	0.107	8.082e-4	2.228	-0.023	2.919e-3
	PN	-3.098	1.114	-3.495e-4	0.166	0.011	-6.739e-4	1.931	-0.020	4.723e-3
	CA	-2.546	1.133	-3.515e-4	0.236	0.090	-1.142e-2	2.177	-0.019	3.746e-3
	NC	-3.012	1.135	-4.491e-4	0.206	0.060	-4.328e-3	2.214	-0.019	3.541e-3

Table 3.8. Mean values of unstandardized model coefficients for hierarchical SIMPLE-BALLNDBH-BA models

SPP.	VAR.	b₀ <i>intercept</i>	b₁ <i>ln(DBH)</i>	b₂ <i>DBH**2</i>	b₃ <i>ln(MAI)</i>	b₄ <i>SL</i>	b₈ <i>EL</i>	b₁₀ <i>CR</i>	b₁₂ <i>BAL/ ln(DBH+1)</i>	b₁₃ <i>BA</i>
RC	WC	-1.810	1.080	-1.643e-4	0.160	-0.071	-1.305e-2	2.492	-0.025	3.418e-3
	PN	-1.853	1.170	-1.381e-4	0.079	0.324	-2.723e-2	2.418	-0.026	2.919e-3
	CA	0.192	1.066	-1.711e-4	-0.197	-0.117	-2.284e-2	2.479	-0.027	2.055e-3
	NC	-1.751	1.072	-1.693e-4	0.217	-0.070	-2.537e-2	2.582	-0.023	2.812e-3
CY	WC	-2.110	1.079	-6.017e-4	0.138	-0.366	-1.586e-2	3.849	-0.032	6.045e-3
	PN	-3.314	1.119	-6.061e-4	0.121	-0.394	-1.175e-2	4.239	-0.030	7.337e-3
	CA	-1.658	0.958	-4.217e-4	0.084	-0.222	-4.807e-3	1.470	-0.030	6.915e-3
	NC	-2.307	0.909	-6.058e-4	0.204	-0.363	-4.555e-3	1.990	-0.030	8.812e-3
TO	WC	-2.980	1.170	-9.080e-4	0.463	0.102	-3.788e-2	2.487	-0.035	9.236e-3
	PN	--	--	--	--	--	--	--	--	--
	CA	-0.497	1.130	-7.185e-4	-0.174	0.134	-1.370e-3	1.034	-0.031	1.067e-2
	NC	-2.651	1.095	-7.933e-4	0.195	0.058	-5.194e-3	2.403	-0.039	9.411e-3
RA	WC	-2.257	0.632	-1.632e-3	0.414	-0.097	-6.734e-3	3.229	-0.025	4.373e-3
	PN	0.938	0.087	-7.089e-4	0.120	-0.059	4.400e-3	2.225	-0.049	7.161e-3
	CA	-0.363	0.470	-1.526e-3	0.163	-0.073	-1.912e-2	3.462	-0.029	3.361e-3
	NC	-0.257	0.335	-1.071e-3	0.146	0.032	-1.695e-2	3.445	-0.015	5.501e-3
BO	WC	-1.544	1.253	-1.753e-4	-0.004	0.168	-1.048e-2	1.569	-0.017	4.876e-4
	PN	-2.460	1.251	-3.340e-4	0.018	-0.016	-1.134e-2	2.954	-0.017	6.715e-4
	CA	-1.412	1.136	-1.996e-4	0.018	0.286	-8.967e-3	1.323	-0.018	-1.392e-3
	NC	-2.123	1.382	-3.864e-4	0.001	-0.016	-7.116e-3	1.413	-0.016	4.523e-3
MA	WC	-2.027	0.957	-8.284e-4	0.103	-0.370	7.316e-3	2.857	-0.035	8.030e-3
	PN	-2.342	1.054	-7.988e-4	0.122	-0.297	-1.588e-3	3.078	-0.035	8.788e-3
	CA	-3.169	1.032	-7.907e-4	0.376	-0.455	5.554e-4	2.812	-0.033	8.714e-3
	NC	-1.949	1.067	-4.633e-4	0.107	-0.414	-1.156e-2	2.680	-0.034	7.557e-3

Table 3.8. Mean values of unstandardized model coefficients for hierarchical SIMPLE-BALLNDBH-BA models

SPP.	VAR.	b ₀ <i>intercept</i>	b ₁ <i>ln(DBH)</i>	b ₂ <i>DBH**2</i>	b ₃ <i>ln(MAI)</i>	b ₄ <i>SL</i>	b ₈ <i>EL</i>	b ₁₀ <i>CR</i>	b ₁₂ <i>BAL/ ln(DBH+1)</i>	b ₁₃ <i>BA</i>
RW	WC	--	--	--	--	--	--	--	--	--
	PN	--	--	--	--	--	--	--	--	--
	CA	--	--	--	--	--	--	--	--	--
	NC	-4.598	1.278	-2.169e-4	0.547	0.051	1.464e-2	2.444	-0.025	2.168e-3
BM	WC	-0.446	0.769	-7.646e-4	0.105	-0.110	-2.467e-2	2.340	-0.040	6.512e-3
	PN	-0.739	0.827	-6.558e-4	0.155	-0.070	-3.058e-2	2.504	-0.045	7.885e-3
	CA	-0.349	0.823	-7.831e-4	-0.046	-0.092	-1.164e-2	1.219	-0.036	7.067e-3
	NC	-1.997	0.958	-7.002e-4	0.104	0.079	-1.979e-2	2.396	-0.035	1.063e-2
NF	WC	-1.906	1.351	-4.765e-4	0.153	0.003	-1.311e-2	2.336	-0.022	2.293e-3
	PN	-3.883	1.301	-4.468e-4	0.483	0.133	-1.808e-2	2.619	-0.002	4.217e-3
	CA	-3.224	1.266	-4.515e-4	0.379	0.088	-1.846e-2	2.633	-0.008	3.184e-3
	NC	-3.143	1.270	-4.592e-4	0.375	0.035	-1.654e-2	2.611	-0.006	3.168e-3

Notes: The values in this table are the mean values of the posterior distribution for model coefficients based on un-standardized covariates. That is, these variables may be used in a Wykoff equation with covariates retained in their original scale and units. These values are point estimates that could be hard-coded into existing FVS source code to update these parameters in the diameter growth models for these 15 species, with the exception that FVS often uses Site Index instead of MAI to estimate site productivity and there is not a clear crosswalk process described to translate between different site species and base ages in FVS. "--" indicates regions where fewer than 10 observations of a tree species existed. Values for b₂, b₈, and b₁₃ are reported using scientific notation where 1.00e-4 represents 1.00x10⁻⁴.

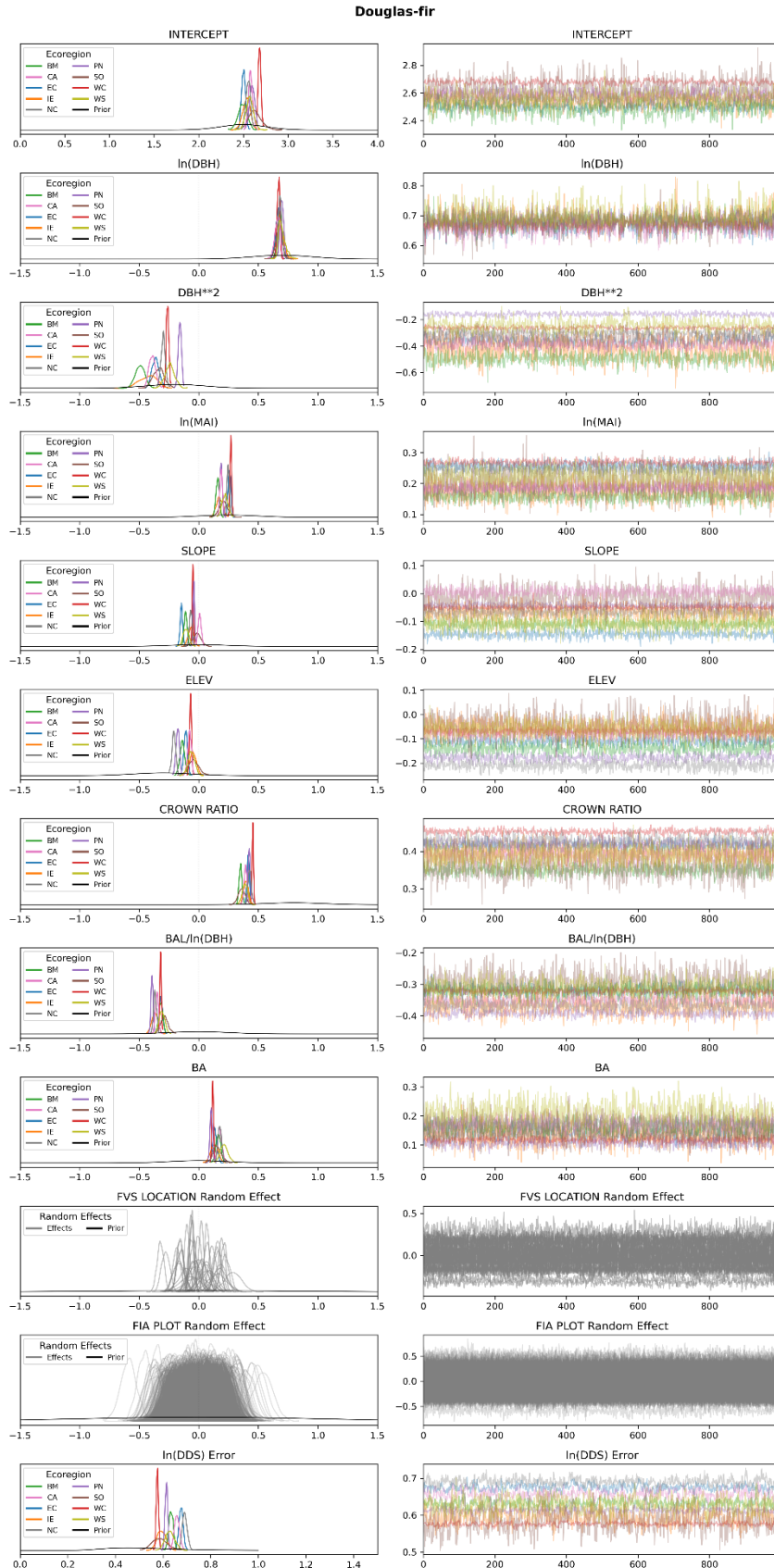


Figure 3.11. Posterior distributions of parameters for the hierarchical benchmark model for Douglas-fir. Posterior distributions for each parameter of the partially pooled SIMPLE-BALLNDBH-BA model are displayed in two ways: kernel density plots of the probability density for the parameter values are shown in the left column of plots; and trace plots showing the values from the parameter over 1,000 steps of the MCMC sampler in the right column of plots. Colors are employed in kernel density plots and trace plots to distinguish the parameters fit for each FVS Variant (shown as “Ecoregion” in the figure legends). The prior distribution used for each parameter is shown as a black line in the kernel density plots. The means and standard deviations of each parameter visualized here are reported in Table 3.7. Parameter values displayed are for standardized covariates to allow the sign and magnitude of regression coefficients to be compared.

Table 3.9. Standard deviations of varying intercepts and random effects for the hierarchical SIMPLE-BALLNDBH-BA model

Common Name	b_0	ϵ_{loc}	ϵ_{plot}
Douglas-fir	0.067 ± 0.01	0.139 ± 0.01	0.122 ± 0.00
Ponderosa pine	0.070 ± 0.03	0.126 ± 0.01	0.121 ± 0.00
Western hemlock	0.084 ± 0.06	0.096 ± 0.01	0.112 ± 0.00
White fir	0.084 ± 0.04	0.108 ± 0.01	0.112 ± 0.00
Pacific silver fir	0.260 ± 0.09	0.088 ± 0.02	0.114 ± 0.00
Mountain hemlock	0.400 ± 0.08	0.109 ± 0.01	0.112 ± 0.00
Western redcedar	0.203 ± 0.06	0.104 ± 0.02	0.105 ± 0.00
Canyon live oak	0.334 ± 0.15	0.133 ± 0.01	0.109 ± 0.00
Tanoak	0.301 ± 0.11	0.105 ± 0.02	0.108 ± 0.00
Red alder	0.139 ± 0.06	0.098 ± 0.01	0.105 ± 0.00
California black oak	0.223 ± 0.12	0.116 ± 0.02	0.105 ± 0.00
Pacific madrone	0.227 ± 0.08	0.104 ± 0.01	0.104 ± 0.00
Coast redwood	0.000 ± 0.00	0.103 ± 0.03	0.122 ± 0.01
Bigleaf maple	0.388 ± 0.07	0.109 ± 0.01	0.102 ± 0.00
Noble fir	0.192 ± 0.09	0.106 ± 0.02	0.108 ± 0.00

Notes: Values in each cell indicate the variation among the geographic units represented by each parameter. These values reflect the standard deviation of the intercept (ecoregion- or FVS Variant-level), FVS Location-level, and FIA Plot-level effects, and are calculated from the posterior distribution for each of these parameters, reporting the average and standard deviation of the among-unit standard deviations.

3.4 DISCUSSION

The observation of consistent biases of current FVS parameterizations overpredicting diameter growth across all species and regions investigated in this study is a significant concern. This type of bias can easily feed opportunistic forest managers and policymakers to overstate the potential for timber volume and value growth, for carbon sequestration or climate mitigation capacity. Based on previous experience manipulating FVS multipliers governing both growth and mortality (Diaz et al., 2018), I also suspect a systematic bias to underpredict mortality may also be contributing to reports of systematic bias in net growth over time (Diaz et al., 2018; Bagdon et al., 2021; Herbert et al., 2023). Given the wide use of FVS to prepare forest management and climate policies, and its heavy use in carbon offset programs, model recalibration to reduce systematic biases should be a high priority for FVS research and development.

The adoption of a decision-rule such as the HDI-ROPE that recognizes practical equivalence in model performance offers a clear departure from a common pattern of Frequentist and Bayesian model comparison approaches that select models based on individual information criteria or point estimates. Had I selected individual models for each species and region based on a single point estimate of a performance criterion such as AIC (Akaike, 1974), DIC (Spiegelhalter et al., 2002, 2014), WAIC, or ELPD-LOO (Vehtari et al., 2017), I would have fallen back into a pattern of proliferating alternative model forms that currently typifies FVS equations across species and regions. In this context, the application of the HDI-ROPE approach clearly indicates an opportunity to significantly consolidate and simplify the variety of models currently being used across Oregon, Washington, and California without sacrificing any practically meaningful predictive performance.

Similar to previous sensitivity analysis by Vacchiano et al. (2008), model fits in this study indicated the tree size terms, crown ratio, and indicators of site productivity are the most influential parameters in the Wykoff process model. Terms related to physical topography and elevation were generally the least influential variables. The use of $BAL/\ln(DBH + 1)$ as the tree-level competition variable consistently showed a strong negative effect on growth, consistent with ecological intuition regarding competition constraining access of each individual to limited available resources. The consistent occurrence of a positive sign on the effect of the stand- or plot-level competition variable (BA) across all regions and species is harder to explain. In general, I would intuitively expect stand-level competition to be negatively related to individual tree diameter growth. The observation that tree- and stand-level competition variables were generally correlated suggests that the model may be overparameterized, and that a single competition term or some decomposition of competition terms may produce model behavior more consistent with ecological principles.

The FIA dataset provided an incredibly valuable proving ground for building and testing models. More elaborate approaches to cross validation and generalization could easily be added to evaluate current and future modifications to FVS models. The preformatting of FIA data into FVS-ready formats at subplot, plot, and condition levels of analysis will be particularly useful for refitting and calibrating additional FVS models over time. Although it was not tested specifically in this study, I expect that the specification of measurement error for both diameter tape measurements and radial increment measurements plays a large influence on the identifiability of

the process model parameters. The remeasurement of diameters at a 10-year interval, and the relatively high measurement error assumed for radial increment measurements are likely to constrain the ability to fit annualized version of the diameter growth model.

Several challenges exist in consistently estimating site productivity. Although Site Index was eschewed from the original Wykoff model for estimating diameter growth because it was considered unreliable for mixed-species and multi-age stands common in the northern Rocky Mountains, Site Index is widely used in equations for diameter growth predictions across western FVS Variants. While FVS allows users to input Site Index values for any species, it is not clearly documented how FVS translates Site Index values for one species into another (e.g., how productivity is estimated for Western hemlock trees when Site Index is based on measurements of Douglas-fir site trees). To mitigate these challenge in model-fitting, I opted to convert Site Index values into a consistent set of units in terms of potential cubic volume increment. However, site productivity is most commonly measured in the field using Site Index, and the conversion of Site Index to potential cubic volume increment relies upon a series of equations and yield curves. In general, measures of site productivity commonly combine climatic, edaphic, and topographic influences on forest growth and mortality into an integrative indicator. The utility of these indicators, which were generally derived for single-species and even-aged forest conditions is uncertain in a changing climate, as the underlying equations and yield curves implicitly assume some climatic normal period. As further developments into FVS are made to consider the addition of climatic drivers, the explanatory ability of site productivity variables should be expected to change, evolving to reflect soil qualities related to nutrient availability or water-holding capacity that are not captured in climatic variables. As climatic drivers are added to the model, it may be preferable to replace proxies for site productivity like MAI or Site Index with field data related more closely to soil qualities.

In the State Space modeling approach adopted in this study, I only incorporated measurement error regarding individual tree measurements of diameter and radial increment. A more robust accounting for sampling uncertainty would also recognize error in other measured predictor variables, such as crown ratio, Site Index, and competition indicators. At the subplot-level, competition variables are likely to be very sensitive to edge effects related to whether individual trees near the plot edge are counted in or out of the plot. Furthermore, the scale at which I expect competition to be affecting tree growth is more likely to be on the order of a tree height,

which can commonly extend more than 5-10 times the reach of a subplot. In these circumstances, I suspect the calculation of competition variables at a larger extent, such as across a whole FIA plot or condition may be more informative estimates for the competitive environment a tree experiences and would reduce the variation in the parameters for competition terms in the model. This approach to utilize measurements from adjacent subplots to quantify competition variables may justify limiting model calibration and refitting to observations of trees on the central FIA subplot (see Figure 3.4).

Although the model form proposed here could be applied using an annualized timestep, the choice to incorporate measurement error and allow for the propagation of process error of the Wykoff model at each timestep limits the ability to retrieve annual estimates of growth factors from measurements made once every ten years given the expected scale of process and measurement errors. I suspect that the use of tree cores collected and processed using more precise dendrochronological analysis along with diameter measurements collected more frequently than once-per-decade may be required to confidently infer parameters for annual growth models that meaningfully reflect and propagate effects of measurement errors and process model uncertainty.

3.5 FUTURE STUDY

The adoption of a hierarchical Bayesian State Space model opens the door to several opportunities for enhancing the FVS model. First, the adoption of multiple observation models for diameter tape and increment measurements can easily be applied to high-precision measurements of diameter growth, which will become increasingly available as laboratory analyses of tree cores that have been collected on FIA plots are released. Precisely-measured annual ring width data can substantially improve the ability of diameter growth models to be fit at an annual timestep. The adoption of a hierarchical model form will allow this type of model to learn from climatic and non-climatic drivers of diameter growth across multiple regions as well, decreasing the current experience of discrete jumps in FVS behavior when ecoregion boundaries are crossed. A hierarchical modeling approach may also enable a significant consolidation of regional FVS variants to cover much broader geographic areas without compromising the predictive performance of the model in each of the regions covered.

As we start to anticipate the application of growth-and-yield in the context of a shifting climate, we expect to see a growing occurrence of climatic conditions that are rare or novel in the

training data if we limit a model's view to a single ecoregion. Considering the response of a tree species to climatic drivers are expected to be shaped substantially by genetic and physiological attributes that are shared to varying degrees among subpopulations across a species' distribution, I suggest it is preferable to establish a model framework which does not assume fixed effects on growth are independent in each ecoregion where a tree species is observed. We may be particularly interested, for example, to consider how tree species in North America respond to climatic changes moving towards hotter and drier conditions by learning from occurrences of that species in hotter and drier conditions further to the south.

Prior to additional data collection and use however, it would be very informative to evaluate the model form using simulated data to identify how much the model is able to retrieve parameter values as the degree of measurement error for each variable increases. The conduct of a simulation study to understand model performance on simulated data is part of emerging best practices in Bayesian workflows (Gelman et al., 2020). Similarly, in addition to simulated data experiments, I would also recommend exploration of more informative prior distributions for model parameters, which are likely necessary to overcome more computationally challenging MCMC hurdles I encountered when using error models with a Laplace distribution, for example.

Although the implementation of a fully Bayesian growth-and-yield framework is not unprecedented (Wilson et al., 2019), I expect incremental progress to be possible and a more likely path for the future development of the FVS model. An intermediate step to substantially increase the reflection of model uncertainty in FVS projections could be made, for example, by storing posterior samples of parameters for each process model, and based on a random seed, selecting which draw from that posterior distribution to use for all parameters. This would side-step the need to directly implement FVS models as random variables to enable stochastic forecasting, but not yet take advantage of the full potential for data assimilation, which I suspect is further into the future for a model like FVS.

3.6 CONCLUSION

The current variety of diameter growth equations across species and regional variants in FVS does not appear to be justified by any practical improvement in predictive performance compared to the adoption of a single version of the Wykoff diameter growth model that drops several of the terms currently being used in FVS. I would strongly encourage the simplification

and consolidation of FVS diameter growth models across western FVS Variants, and the consideration of a multi-regional version of the model covering a larger span of existing FVS Variants. This need for FVS recalibration is also clearly motivated by a growing number of studies finding systematic bias of FVS in several western Variants to overpredict forest growth.

The FIA dataset offers an unprecedented proving ground to consistently refit and evaluate FVS's predictive performance against diverse forest conditions in transparent, reproducible, and unbiased way. I encourage the development of a turn-key recalibration workflow for updating FVS component models as new FIA data are released each year, and reporting of FVS model performance against FIA data as part of the documentation and release of new versions of the software. The State Space modeling framework, model comparison approach, and Bayesian data analysis more generally offer substantial opportunities to enhance the Forest Vegetation Simulator and to increase transparency and confidence in the development, maintenance, and application of this software to support improved forest management and conservation in the United States and beyond.

3.7 SUPPLEMENTARY MATERIALS

The Forest Carbon Modeling Group (FCMG) recently emerged as a working group organized by researchers from the US Forest Service (USFS) and National Council for Air and Stream Improvement to identify a set of research and development priorities for forest carbon modeling. Over the first few months of 2023, the FCMG convened several meetings with 85 participants from academia (23), government (21), industry (36), and non-governmental organizations (5) with diverse expertise and professional experience in forest and carbon modeling. This group has placed particular focus on identifying opportunities to improve the Forest Vegetation Simulator (FVS). Participants were polled to gauge the importance they assigned to 15 different topics related to “Platforms and Accessibility” of the FVS software, rating each topic from a scale of 1 (low priority) to 5 (high priority). Results from polling of FCMG participants were shared via e-mail in a draft report and excerpted here (US Forest Carbon Modeling Group, in preparation).

Table S3.10. Prioritization of FVS research and development topics by the Forest Carbon Modeling Group.

Rank	Topic	EDU	GOV	COM	ORG	All
1	Include models that can adequately model management for all forest types and ownerships (e.g., industrial), not just those built on data from federal lands	4.17	4.44	4.18	4.25	4.26
2	Dynamically calibrate FVS with FIA data on growth/mortality rates and/or address differences in growth or mortality across ownerships to help address shortcomings	3.56	4.90	3.91	4.25	4.16
3	Create partnerships with forestry research cooperatives to obtain models for updating and improving FVS	4.00	4.30	3.77	4.25	4.08
4	Analyze FVS models for strengths, weaknesses, opportunities, threats (SWOT)	3.78	4.20	3.64	4.25	3.97
5	Create a more robust propagation of uncertainty to approach total uncertainty	3.89	4.00	3.55	4.25	3.92
6	Improve FVS documentation to enhance accessibility: 1) What should and shouldn't be modified in FVS by variant; and 2) Model parameters and what can be adjusted to better calibrate the model	3.44	4.33	3.41	4.50	3.92
7	Improve tree form, mortality, and regeneration models	3.83	4.70	3.38	3.75	3.92
8	Make FVS platform independent, where code is more broadly applied, understood, and accessible	3.44	3.44	3.18	4.50	3.64
9	Integrate FVS with spatially explicit processes to extend its capabilities	3.56	3.50	2.76	3.50	3.33
10	Determine if FVS could be developed for modeling larger scales (i.e., upscaling; aggregation from stand- to landscape-level).	3.17	3.44	2.64	4.00	3.31
11	Move beyond the "variant" paradigm to a more comprehensive system that is less jurisdictionally dependent	3.33	3.70	3.14	3.00	3.29
12	Include more variables (e.g., belowground carbon, maximum basal area, and site class modeling)	3.33	3.56	2.82	3.00	3.18
13	Incorporate more complex silviculture within a site and adjacent areas	3.67	2.78	3.55	2.25	3.06
14	Offer tiered scales that allow for bringing in local variables	3.17	3.78	2.68	2.50	3.03
15	Separate growth and yield versus carbon accounting/modeling/estimation as separate FVS components	2.72	2.67	3.52	2.00	2.73

Notes: Values in each cell represent the average rating among respondents within each of four groups: academia (EDU); government (GOV); industry (COM); and non-profit (ORG). "All" shows the average with each group receiving equal weight. Table excerpted from an unpublished FCMG draft report (US Forest Carbon Modeling Group, in preparation).

Fitting a Chapman-Richards model to impute missing tree heights

As mentioned briefly in the main text, trees without field-observed total heights in the FIA dataset were imputed using a newly fitted regression equation. I used a Chapman-Richards model form to relate tree height from diameter at breast height:

$$\hat{H} = 4.5 + b_0 * (1 - e^{-b_1 D})^{b_2} \quad (3.13)$$

where $\hat{H}$ is estimated total tree height, D is diameter at breast height (4.5 feet), b_0 represents an asymptote for the maximum height above breast height for that tree species, and b_1 and b_2 are shape parameters.

Total tree height was defined to follow a normal distribution with error in height predictions increasing with increasing diameter:

$$H \sim N(\hat{H}, \sigma + \gamma D) \quad (3.14)$$

where H is the distribution of predicted tree heights, σ represents the minimum amount of variation in height estimates regardless of tree size, and γ encodes an additional amount of variation that increases linearly with tree diameter (D). An inverse gamma prior was defined for σ as $\sigma \sim IG(1.0, 20.0)$. A uniform prior was defined for γ as $\gamma \sim U(0, 10)$. Each b parameter was constrained to be positive and modeled to

follow a gamma distribution, parameterized by converting estimates of the mean and standard deviation for the distribution into concentration and rate parameters. The model was fit using a hierarchical form, with distinct parameters estimated for each FVS Location, drawn from a shared parent distribution across all FVS Locations. Each b parameter for each FVS Location was drawn from a Gamma distribution with mean and standard deviation as displayed in Table S3.37.

Table S3.11. Hierarchical parameterization of Chapman-Richards height-diameter model

Parameter	Hierarchical Priors	
	mean	standard deviation
b_0	G(200, 150)	IG(1.0, 20.0)
b_1	G(0.1, 0.1)	IG(0.1, 0.1)
b_2	G(1.5, 1.0)	IG(1.0, 3.0)

Note: parameters given above for Gamma (G) distributions are defined using mean and standard deviation, not concentration and rate. Parameters for inverse gamma (IG) distributions are concentration and rate.

The model for each species was fit using MCMC sampling with three chains using the No U-Turn Sampler (NUTS) with 5,000 burn-in (discarded) samples followed by 1,000 retained

samples. MCMC convergence was confirmed using the Gelman-Rubin R-hat statistic and visually through traceplots and kernel density estimates. Posterior predictive checks were also made to compare how the distributions of predicted heights compare to observed heights.

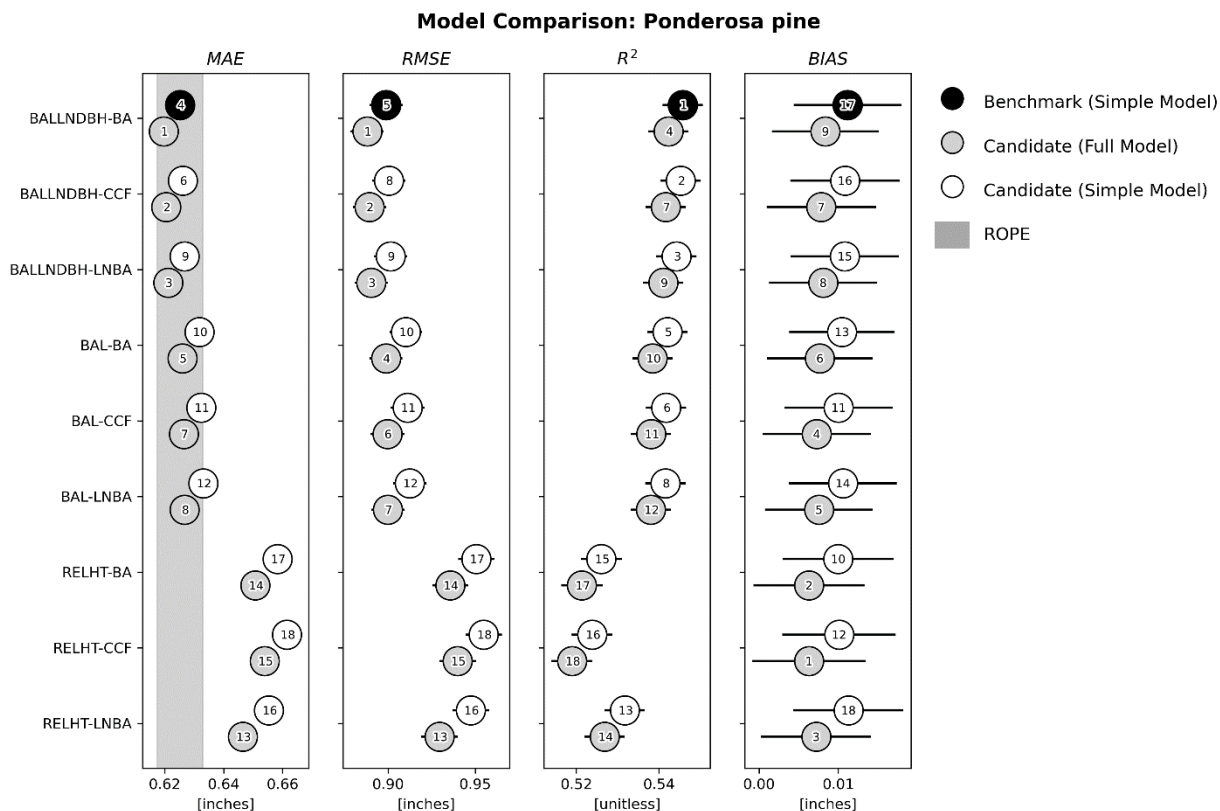


Figure S3.12. Unpooled model comparison with multiple performance metrics: Ponderosa pine.

These graphs illustrate the performance of the unpooled benchmark model (SIMPLE-BALLNDBH-BAL) as well as the 17 other unpooled candidate models using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Bayesian R^2 , and average bias. These performance metrics are calculated from all tree observations across the four focal FVS Variants. The combination of stand-level and tree-level competition variables used in each model are labeled at the left side of the graphs. The number inside each dot indicates the rank of the model based on the model's average for that performance metric across the four focal FVS Variants. The location of each dot along the x-axis indicates the mean value for the performance metric, and horizontal lines (if visible beyond the dot) indicate the standard deviation of the performance metric. The benchmark model is filled in black, while simple-form candidate models are hollow, and full-form candidate models are filled in gray. The vertical gray bar in the MAE graph is intended to illustrate the approximate size of the Region of Practical Equivalence employed in HDI-ROPE tests. However, it is worth noting the HDI-ROPE test was applied in each region independently (as opposed to using a single ROPE across all regions as shown above). The HDI-ROPE test was also applied using the 90% highest posterior density interval, as opposed to the mean $\pm$ the standard deviation.

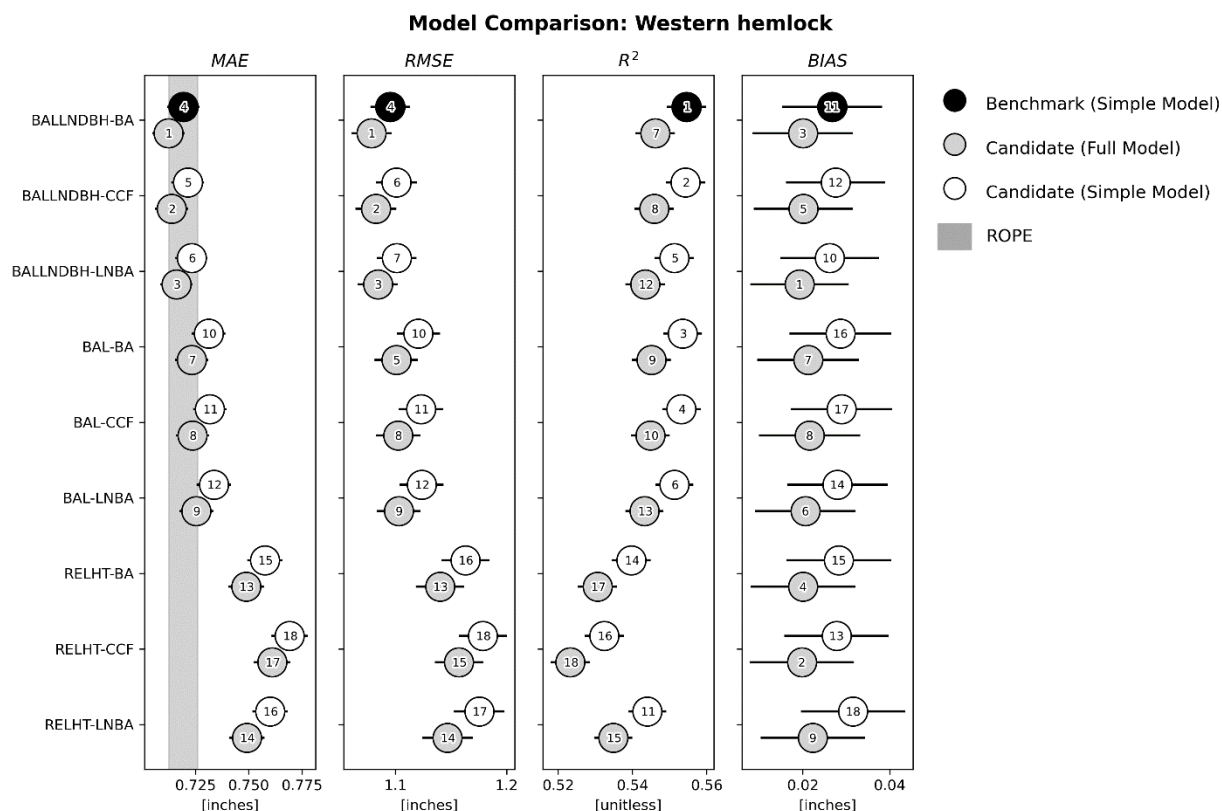


Figure S3.13. Unpooled model comparison with multiple performance metrics: Western hemlock.

These graphs illustrate the performance of the unpooled benchmark model (SIMPLE-BALLNDBH-BAL) as well as the 17 other unpooled candidate models using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Bayesian R^2 , and average bias. These performance metrics are calculated from all tree observations across the four focal FVS Variants. The combination of stand-level and tree-level competition variables used in each model are labeled at the left side of the graphs. The number inside each dot indicates the rank of the model based on the model's average for that performance metric across the four focal FVS Variants. The location of each dot along the x-axis indicates the mean value for the performance metric, and horizontal lines (if visible beyond the dot) indicate the standard deviation of the performance metric. The benchmark model is filled in black, while simple-form candidate models are hollow, and full-form candidate models are filled in gray. The vertical gray bar in the MAE graph is intended to illustrate the approximate size of the Region of Practical Equivalence employed in HDI-ROPE tests. However, it is worth noting the HDI-ROPE test was applied in each region independently (as opposed to using a single ROPE across all regions as shown above). The HDI-ROPE test was also applied using the 90% highest posterior density interval, as opposed to the mean $\pm$ the standard deviation.

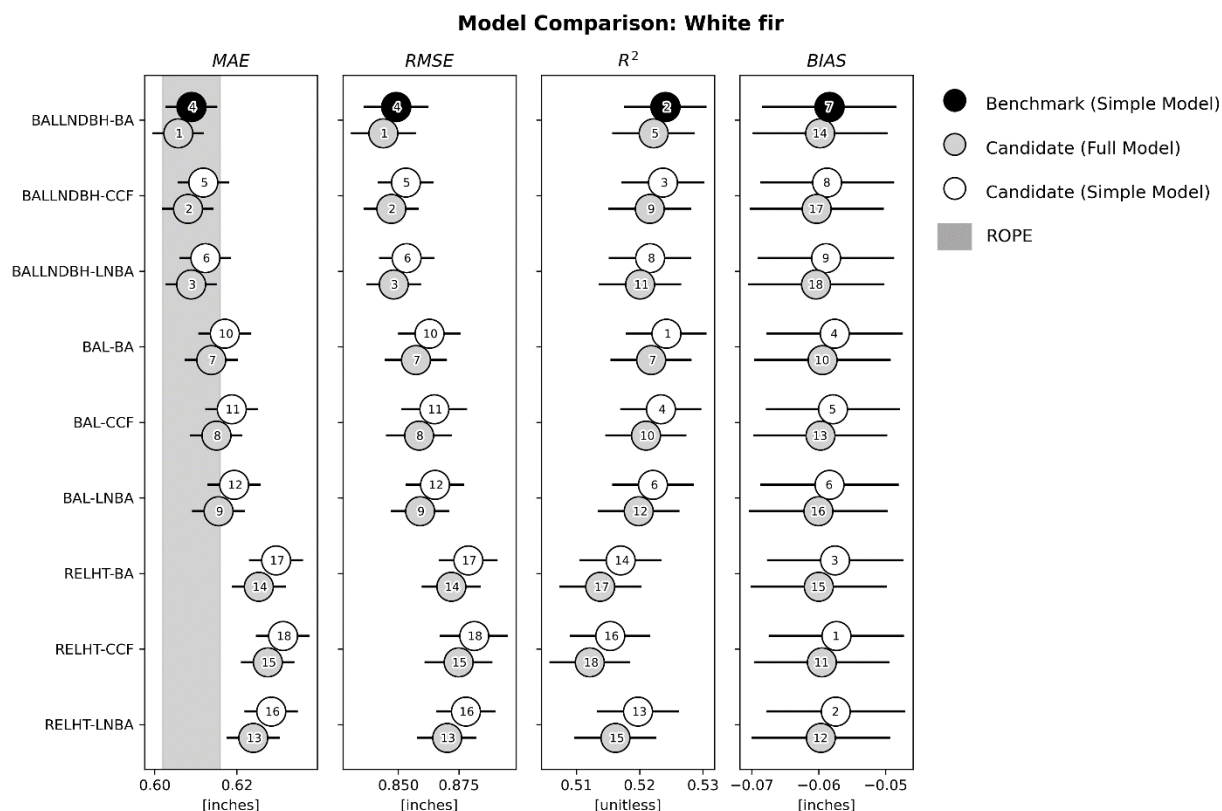


Figure S3.14. Unpooled model comparison with multiple performance metrics: White fir
 These graphs illustrate the performance of the unpooled benchmark model (SIMPLE-BALLNDBH-BAL) as well as the 17 other unpooled candidate models using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Bayesian R^2 , and average bias. These performance metrics are calculated from all tree observations across the four focal FVS Variants. The combination of stand-level and tree-level competition variables used in each model are labeled at the left side of the graphs. The number inside each dot indicates the rank of the model based on the model's average for that performance metric across the four focal FVS Variants. The location of each dot along the x-axis indicates the mean value for the performance metric, and horizontal lines (if visible beyond the dot) indicate the standard deviation of the performance metric. The benchmark model is filled in black, while simple-form candidate models are hollow, and full-form candidate models are filled in gray. The vertical gray bar in the MAE graph is intended to illustrate the approximate size of the Region of Practical Equivalence employed in HDI-ROPE tests. However, it is worth noting the HDI-ROPE test was applied in each region independently (as opposed to using a single ROPE across all regions as shown above). The HDI-ROPE test was also applied using the 90% highest posterior density interval, as opposed to the mean $\pm$ the standard deviation.

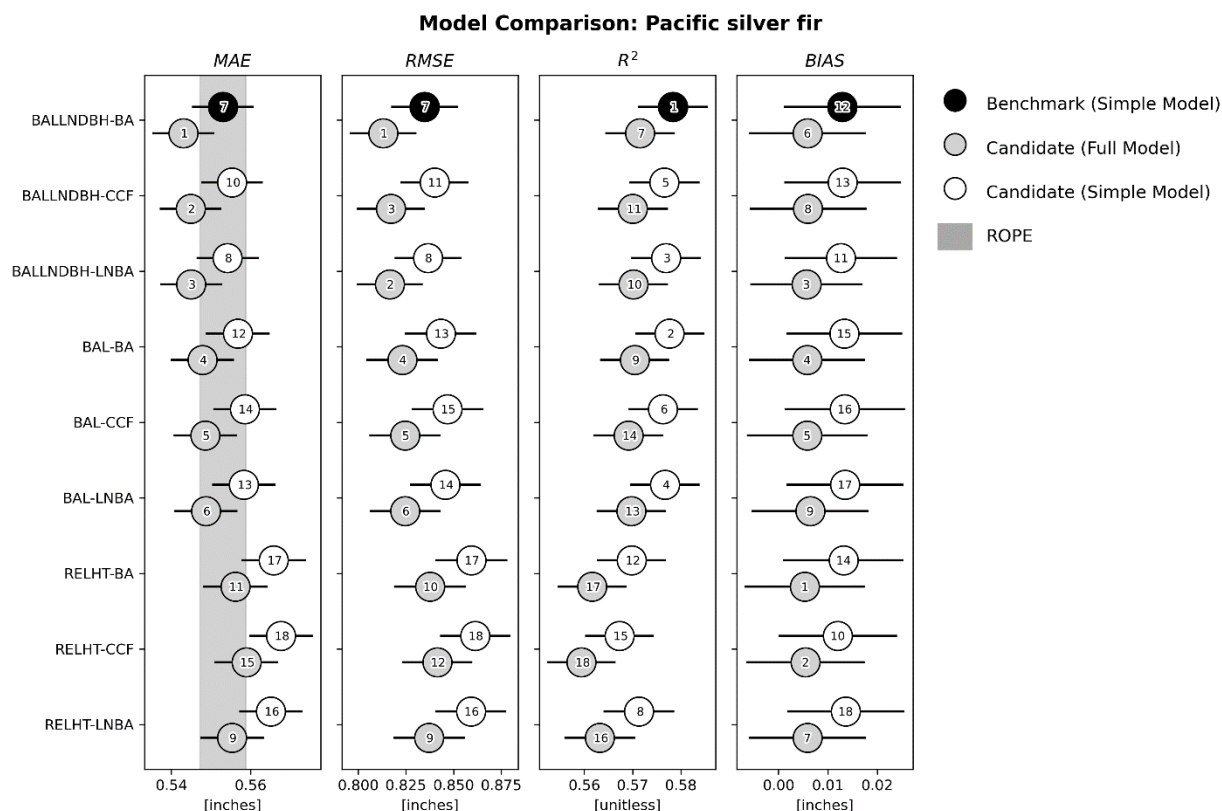


Figure S3.15. Unpooled model comparison with multiple performance metrics: Pacific silver fir.

These graphs illustrate the performance of the unpooled benchmark model (SIMPLE-BALLNDBH-BAL) as well as the 17 other unpooled candidate models using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Bayesian R^2 , and average bias. These performance metrics are calculated from all tree observations across the four focal FVS Variants. The combination of stand-level and tree-level competition variables used in each model are labeled at the left side of the graphs. The number inside each dot indicates the rank of the model based on the model's average for that performance metric across the four focal FVS Variants. The location of each dot along the x-axis indicates the mean value for the performance metric, and horizontal lines (if visible beyond the dot) indicate the standard deviation of the performance metric. The benchmark model is filled in black, while simple-form candidate models are hollow, and full-form candidate models are filled in gray. The vertical gray bar in the MAE graph is intended to illustrate the approximate size of the Region of Practical Equivalence employed in HDI-ROPE tests. However, it is worth noting the HDI-ROPE test was applied in each region independently (as opposed to using a single ROPE across all regions as shown above). The HDI-ROPE test was also applied using the 90% highest posterior density interval, as opposed to the mean $\pm$ the standard deviation.

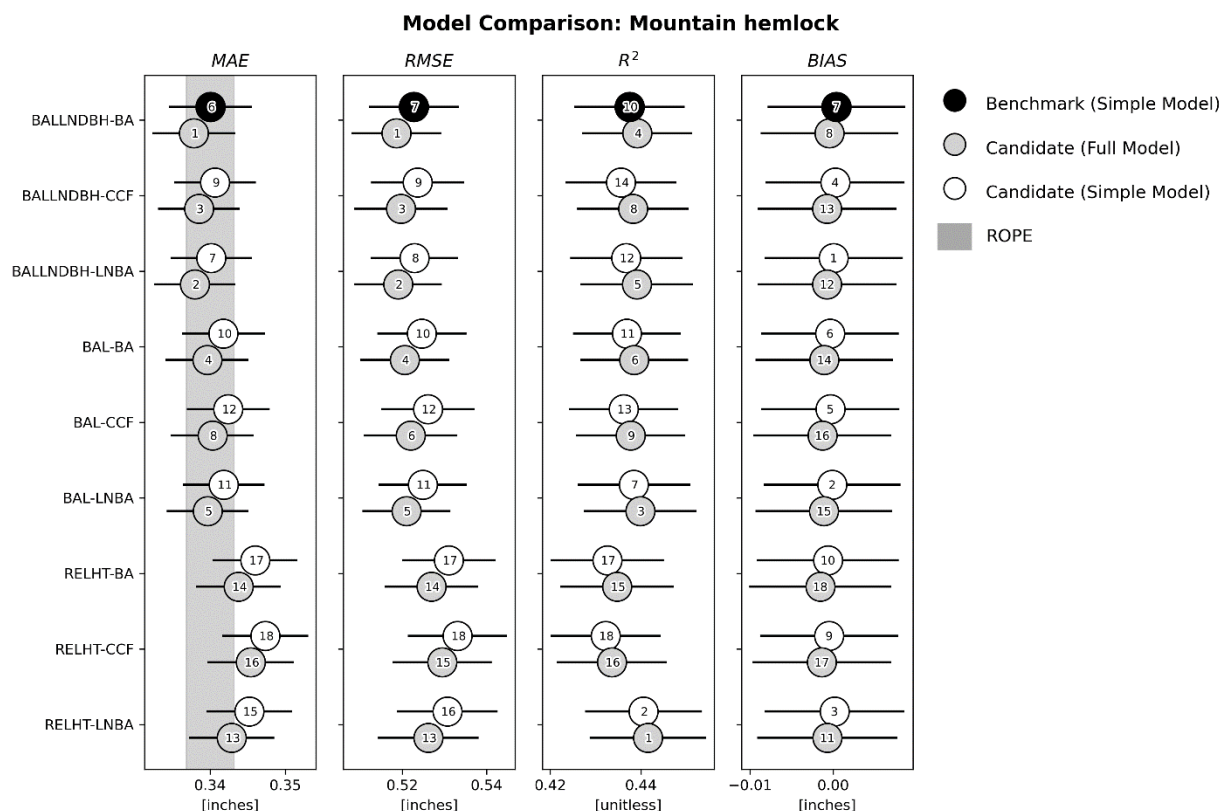


Figure S3.16. Unpooled model comparison with multiple performance metrics: Mountain hemlock.

These graphs illustrate the performance of the unpooled benchmark model (SIMPLE-BALLNDBH-BAL) as well as the 17 other unpooled candidate models using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Bayesian R^2 , and average bias. These performance metrics are calculated from all tree observations across the four focal FVS Variants. The combination of stand-level and tree-level competition variables used in each model are labeled at the left side of the graphs. The number inside each dot indicates the rank of the model based on the model's average for that performance metric across the four focal FVS Variants. The location of each dot along the x-axis indicates the mean value for the performance metric, and horizontal lines (if visible beyond the dot) indicate the standard deviation of the performance metric. The benchmark model is filled in black, while simple-form candidate models are hollow, and full-form candidate models are filled in gray. The vertical gray bar in the MAE graph is intended to illustrate the approximate size of the Region of Practical Equivalence employed in HDI-ROPE tests. However, it is worth noting the HDI-ROPE test was applied in each region independently (as opposed to using a single ROPE across all regions as shown above). The HDI-ROPE test was also applied using the 90% highest posterior density interval, as opposed to the mean $\pm$ the standard deviation.

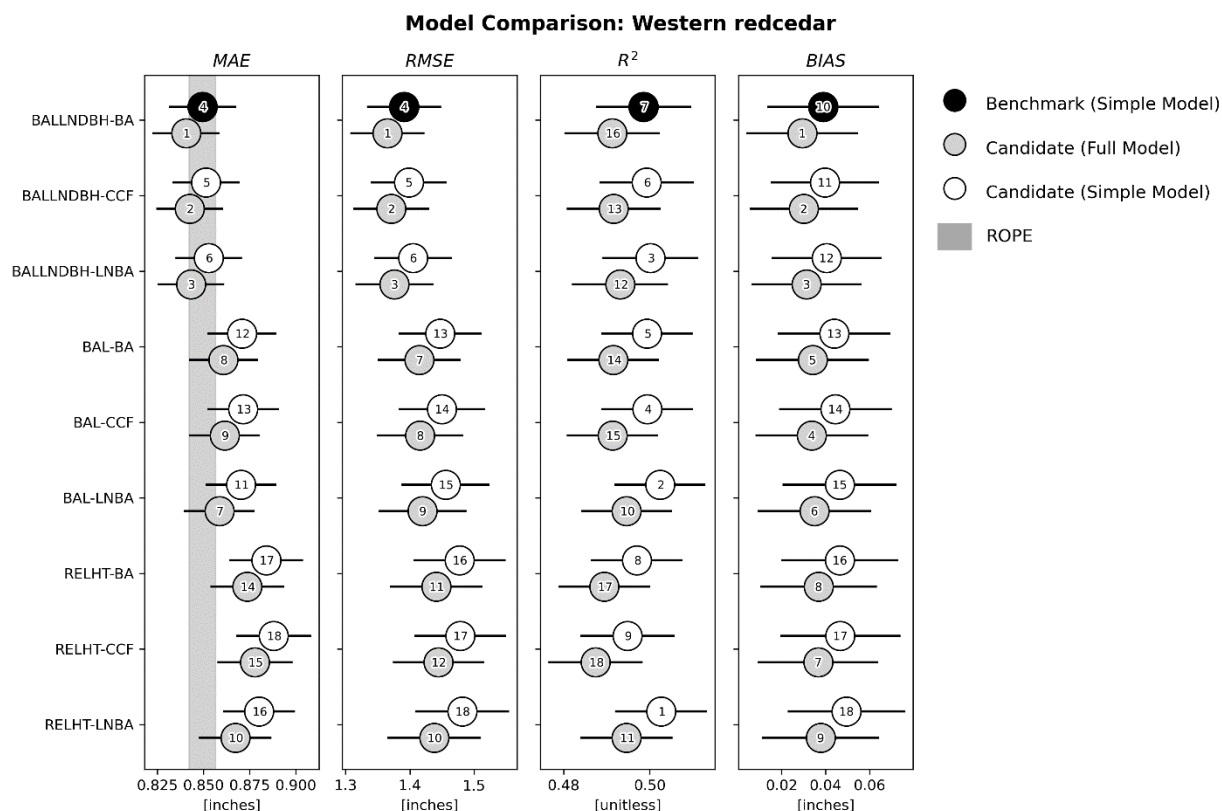


Figure S3.17. Unpooled model comparison with multiple performance metrics: Western redcedar.

These graphs illustrate the performance of the unpooled benchmark model (SIMPLE-BALLNDBH-BAL) as well as the 17 other unpooled candidate models using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Bayesian R^2 , and average bias. These performance metrics are calculated from all tree observations across the four focal FVS Variants. The combination of stand-level and tree-level competition variables used in each model are labeled at the left side of the graphs. The number inside each dot indicates the rank of the model based on the model's average for that performance metric across the four focal FVS Variants. The location of each dot along the x-axis indicates the mean value for the performance metric, and horizontal lines (if visible beyond the dot) indicate the standard deviation of the performance metric. The benchmark model is filled in black, while simple-form candidate models are hollow, and full-form candidate models are filled in gray. The vertical gray bar in the MAE graph is intended to illustrate the approximate size of the Region of Practical Equivalence employed in HDI-ROPE tests. However, it is worth noting the HDI-ROPE test was applied in each region independently (as opposed to using a single ROPE across all regions as shown above). The HDI-ROPE test was also applied using the 90% highest posterior density interval, as opposed to the mean $\pm$ the standard deviation.

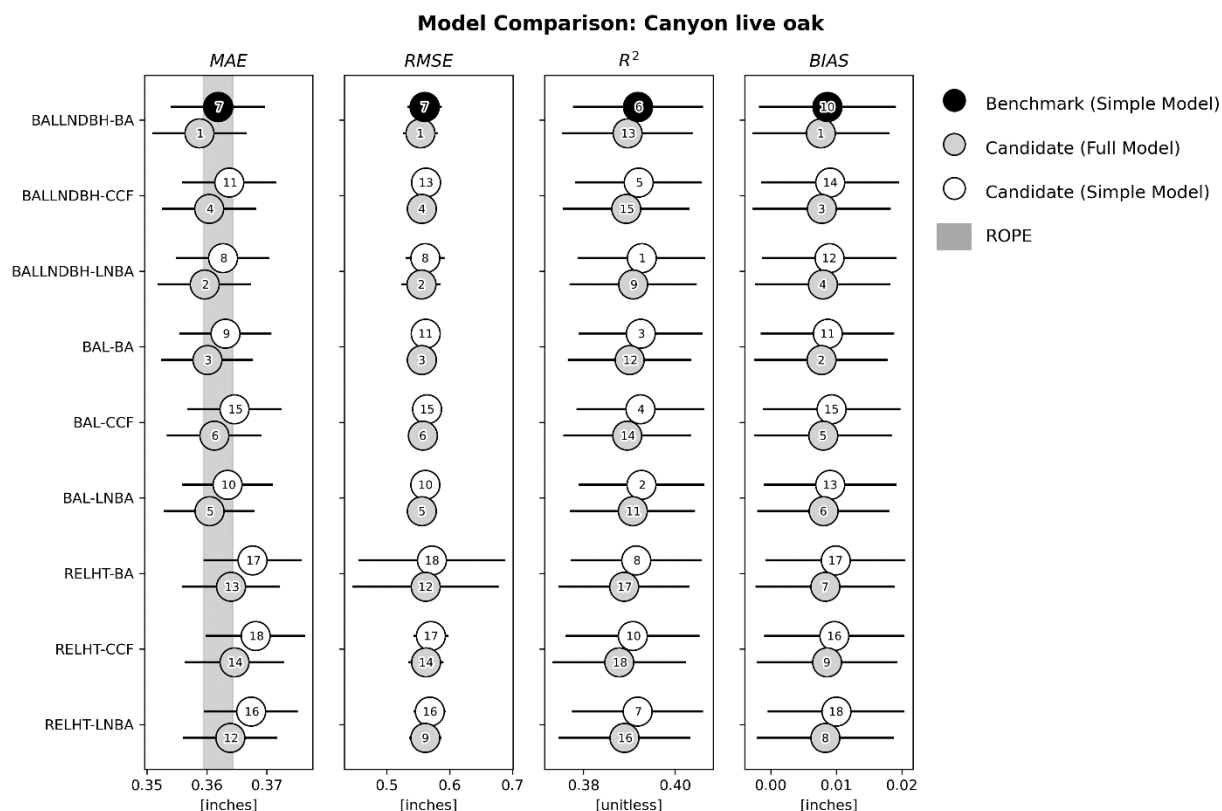


Figure S3.18. Unpooled model comparison with multiple performance metrics: Canyon live oak.

These graphs illustrate the performance of the unpooled benchmark model (SIMPLE-BALLNDBH-BAL) as well as the 17 other unpooled candidate models using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Bayesian R^2 , and average bias. These performance metrics are calculated from all tree observations across the four focal FVS Variants. The combination of stand-level and tree-level competition variables used in each model are labeled at the left side of the graphs. The number inside each dot indicates the rank of the model based on the model's average for that performance metric across the four focal FVS Variants. The location of each dot along the x-axis indicates the mean value for the performance metric, and horizontal lines (if visible beyond the dot) indicate the standard deviation of the performance metric. The benchmark model is filled in black, while simple-form candidate models are hollow, and full-form candidate models are filled in gray. The vertical gray bar in the MAE graph is intended to illustrate the approximate size of the Region of Practical Equivalence employed in HDI-ROPE tests. However, it is worth noting the HDI-ROPE test was applied in each region independently (as opposed to using a single ROPE across all regions as shown above). The HDI-ROPE test was also applied using the 90% highest posterior density interval, as opposed to the mean $\pm$ the standard deviation.

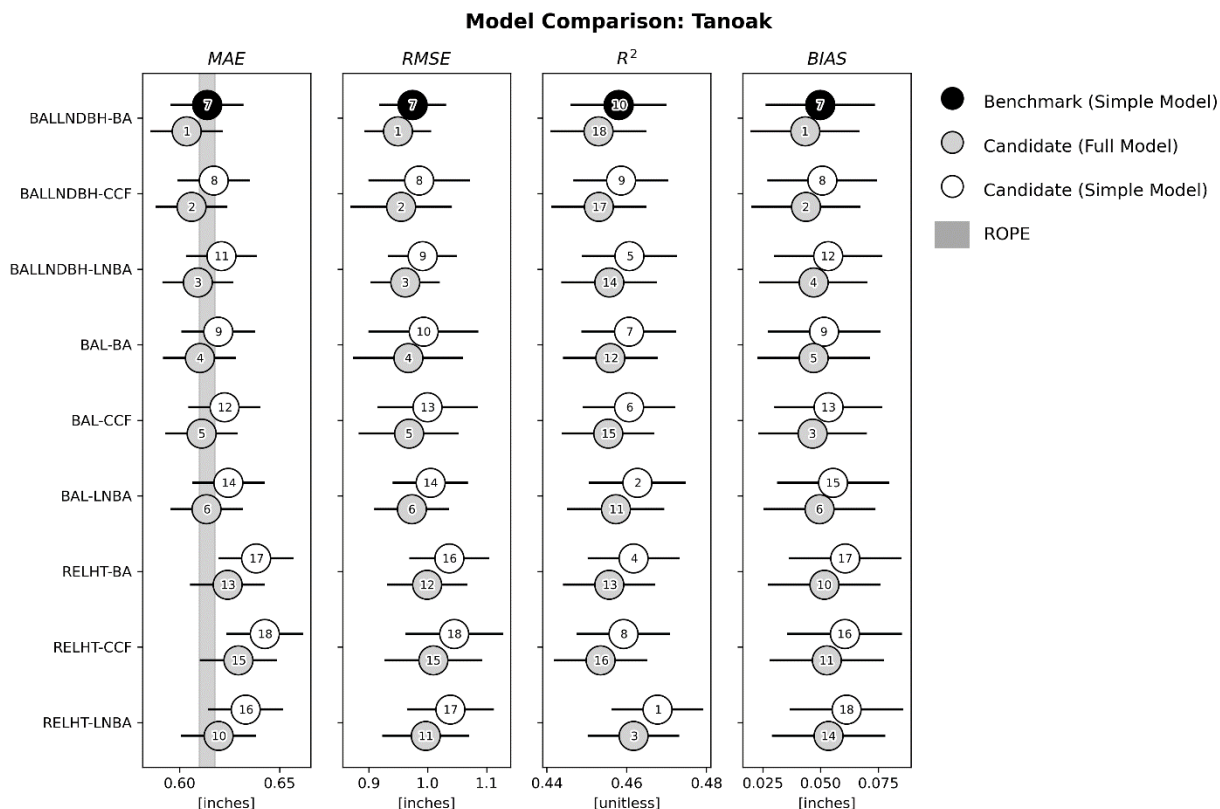


Figure S3.19. Unpooled model comparison with multiple performance metrics: Tanoak. These graphs illustrate the performance of the unpooled benchmark model (SIMPLE-BALLNDBH-BAL) as well as the 17 other unpooled candidate models using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Bayesian R^2 , and average bias. These performance metrics are calculated from all tree observations across the four focal FVS Variants. The combination of stand-level and tree-level competition variables used in each model are labeled at the left side of the graphs. The number inside each dot indicates the rank of the model based on the model's average for that performance metric across the four focal FVS Variants. The location of each dot along the x-axis indicates the mean value for the performance metric, and horizontal lines (if visible beyond the dot) indicate the standard deviation of the performance metric. The benchmark model is filled in black, while simple-form candidate models are hollow, and full-form candidate models are filled in gray. The vertical gray bar in the MAE graph is intended to illustrate the approximate size of the Region of Practical Equivalence employed in HDI-ROPE tests. However, it is worth noting the HDI-ROPE test was applied in each region independently (as opposed to using a single ROPE across all regions as shown above). The HDI-ROPE test was also applied using the 90% highest posterior density interval, as opposed to the mean $\pm$ the standard deviation.

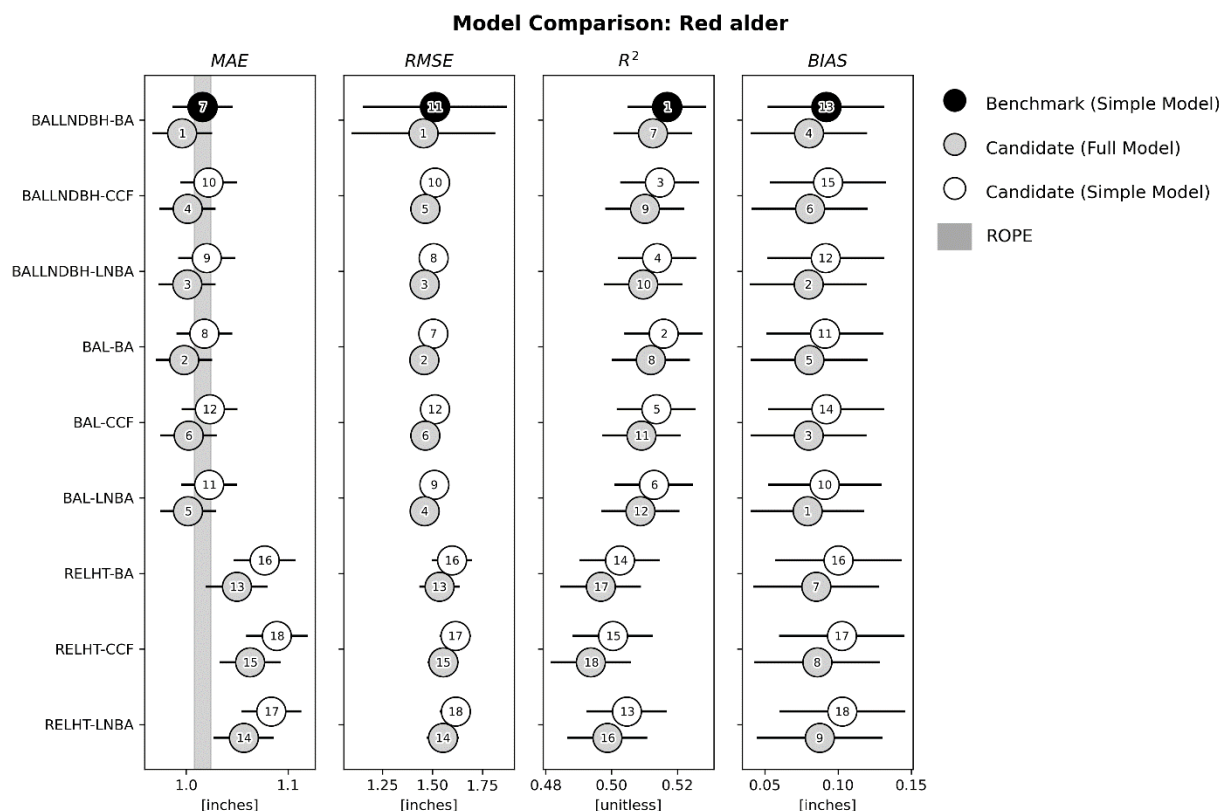


Figure S3.20. Unpooled model comparison with multiple performance metrics: Red alder. These graphs illustrate the performance of the unpooled benchmark model (SIMPLE-BALLNDBH-BAL) as well as the 17 other unpooled candidate models using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Bayesian R^2 , and average bias. These performance metrics are calculated from all tree observations across the four focal FVS Variants. The combination of stand-level and tree-level competition variables used in each model are labeled at the left side of the graphs. The number inside each dot indicates the rank of the model based on the model's average for that performance metric across the four focal FVS Variants. The location of each dot along the x-axis indicates the mean value for the performance metric, and horizontal lines (if visible beyond the dot) indicate the standard deviation of the performance metric. The benchmark model is filled in black, while simple-form candidate models are hollow, and full-form candidate models are filled in gray. The vertical gray bar in the MAE graph is intended to illustrate the approximate size of the Region of Practical Equivalence employed in HDI-ROPE tests. However, it is worth noting the HDI-ROPE test was applied in each region independently (as opposed to using a single ROPE across all regions as shown above). The HDI-ROPE test was also applied using the 90% highest posterior density interval, as opposed to the mean $\pm$ the standard deviation.

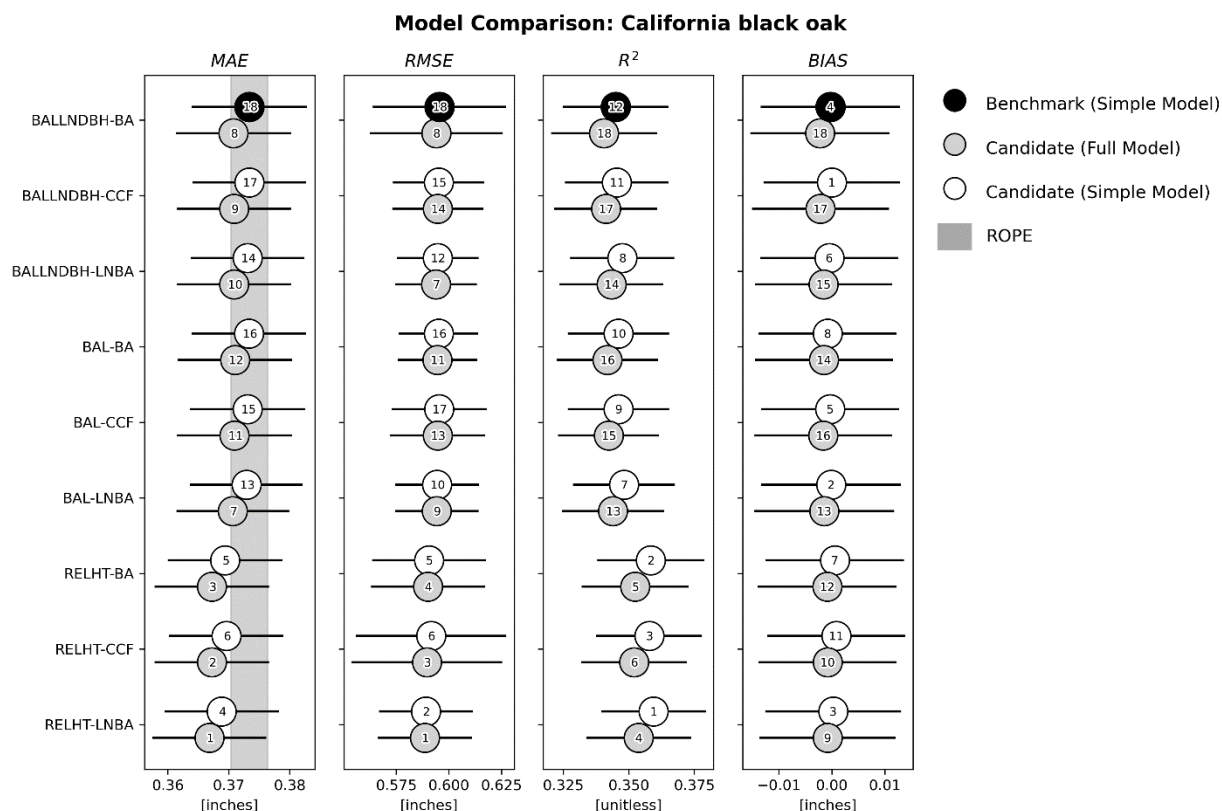


Figure S3.21. Unpooled model comparison with multiple performance metrics: California black oak.

These graphs illustrate the performance of the unpooled benchmark model (SIMPLE-BALLNDBH-BAL) as well as the 17 other unpooled candidate models using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Bayesian R^2 , and average bias. These performance metrics are calculated from all tree observations across the four focal FVS Variants. The combination of stand-level and tree-level competition variables used in each model are labeled at the left side of the graphs. The number inside each dot indicates the rank of the model based on the model's average for that performance metric across the four focal FVS Variants. The location of each dot along the x-axis indicates the mean value for the performance metric, and horizontal lines (if visible beyond the dot) indicate the standard deviation of the performance metric. The benchmark model is filled in black, while simple-form candidate models are hollow, and full-form candidate models are filled in gray. The vertical gray bar in the MAE graph is intended to illustrate the approximate size of the Region of Practical Equivalence employed in HDI-ROPE tests. However, it is worth noting the HDI-ROPE test was applied in each region independently (as opposed to using a single ROPE across all regions as shown above). The HDI-ROPE test was also applied using the 90% highest posterior density interval, as opposed to the mean $\pm$ the standard deviation.

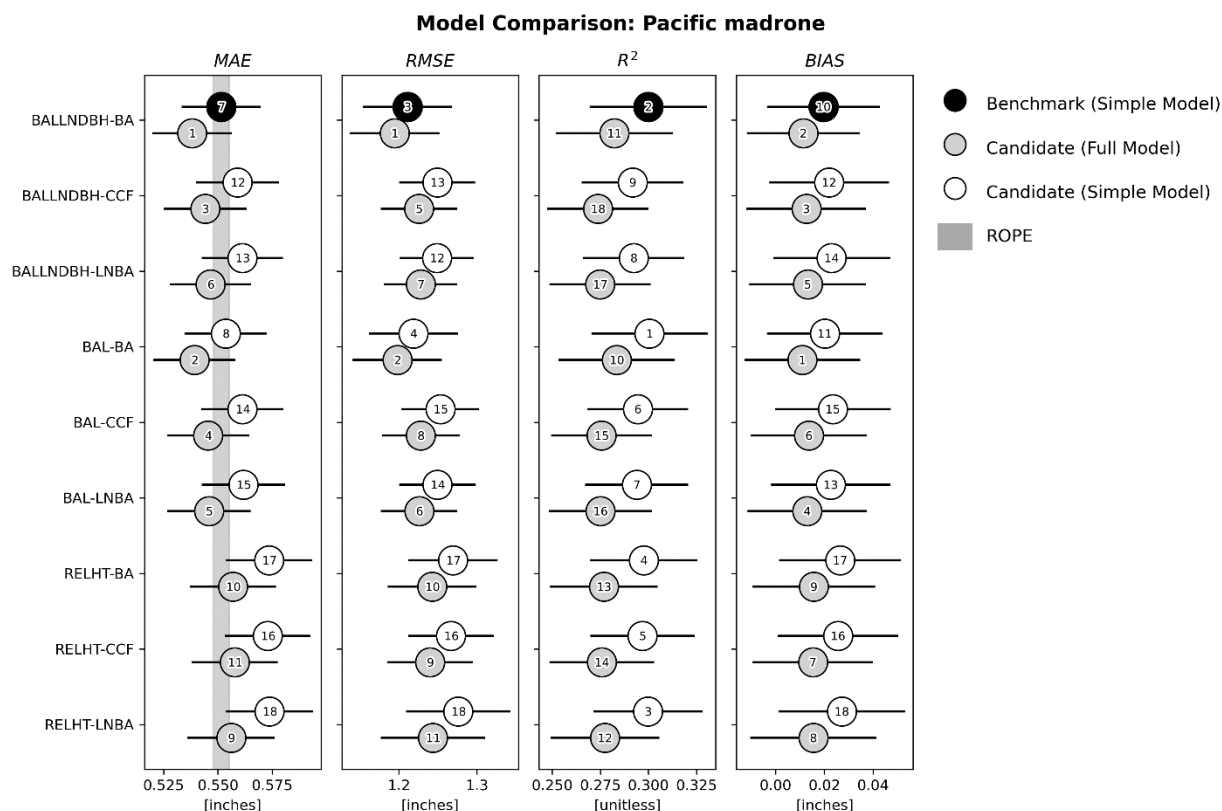


Figure S3.22. Unpooled model comparison with multiple performance metrics: Pacific madrone.

These graphs illustrate the performance of the unpooled benchmark model (SIMPLE-BALLNDBH-BAL) as well as the 17 other unpooled candidate models using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Bayesian R^2 , and average bias. These performance metrics are calculated from all tree observations across the four focal FVS Variants. The combination of stand-level and tree-level competition variables used in each model are labeled at the left side of the graphs. The number inside each dot indicates the rank of the model based on the model's average for that performance metric across the four focal FVS Variants. The location of each dot along the x-axis indicates the mean value for the performance metric, and horizontal lines (if visible beyond the dot) indicate the standard deviation of the performance metric. The benchmark model is filled in black, while simple-form candidate models are hollow, and full-form candidate models are filled in gray. The vertical gray bar in the MAE graph is intended to illustrate the approximate size of the Region of Practical Equivalence employed in HDI-ROPE tests. However, it is worth noting the HDI-ROPE test was applied in each region independently (as opposed to using a single ROPE across all regions as shown above). The HDI-ROPE test was also applied using the 90% highest posterior density interval, as opposed to the mean $\pm$ the standard deviation.

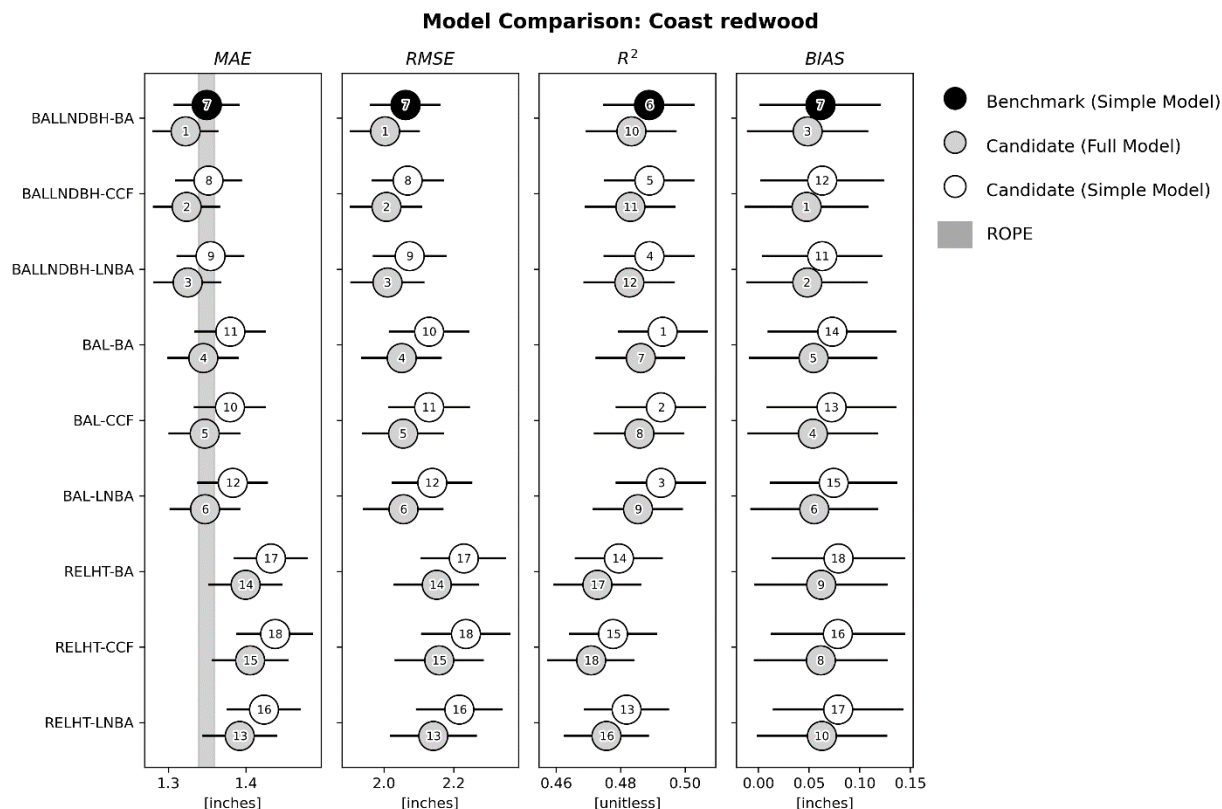


Figure S3.23. Unpooled model comparison with multiple performance metrics: Coast redwood.

These graphs illustrate the performance of the unpooled benchmark model (SIMPLE-BALLNDBH-BAL) as well as the 17 other unpooled candidate models using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Bayesian R^2 , and average bias. These performance metrics are calculated from all tree observations across the four focal FVS Variants. The combination of stand-level and tree-level competition variables used in each model are labeled at the left side of the graphs. The number inside each dot indicates the rank of the model based on the model's average for that performance metric across the four focal FVS Variants. The location of each dot along the x-axis indicates the mean value for the performance metric, and horizontal lines (if visible beyond the dot) indicate the standard deviation of the performance metric. The benchmark model is filled in black, while simple-form candidate models are hollow, and full-form candidate models are filled in gray. The vertical gray bar in the MAE graph is intended to illustrate the approximate size of the Region of Practical Equivalence employed in HDI-ROPE tests. However, it is worth noting the HDI-ROPE test was applied in each region independently (as opposed to using a single ROPE across all regions as shown above). The HDI-ROPE test was also applied using the 90% highest posterior density interval, as opposed to the mean $\pm$ the standard deviation.

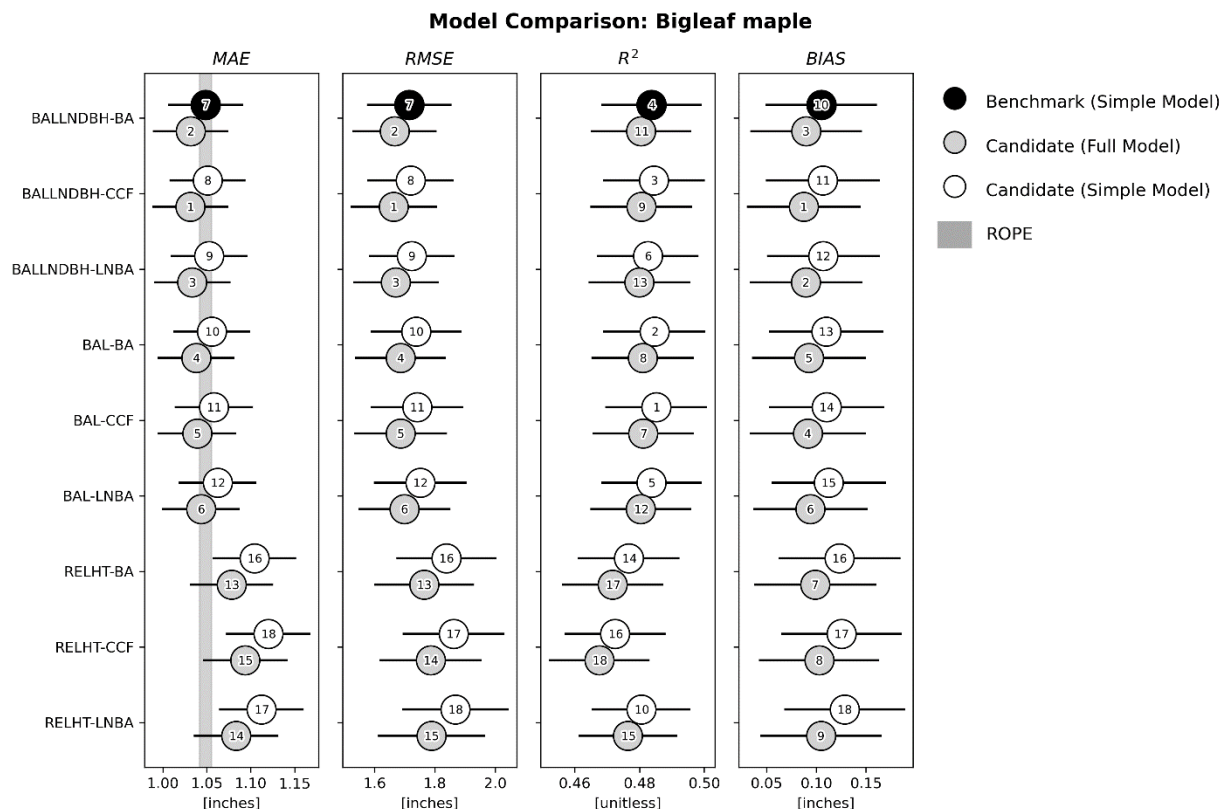


Figure S3.24. Unpooled model comparison with multiple performance metrics: Bigleaf maple.

These graphs illustrate the performance of the unpooled benchmark model (SIMPLE-BALLNDBH-BAL) as well as the 17 other unpooled candidate models using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Bayesian R^2 , and average bias. These performance metrics are calculated from all tree observations across the four focal FVS Variants. The combination of stand-level and tree-level competition variables used in each model are labeled at the left side of the graphs. The number inside each dot indicates the rank of the model based on the model's average for that performance metric across the four focal FVS Variants. The location of each dot along the x-axis indicates the mean value for the performance metric, and horizontal lines (if visible beyond the dot) indicate the standard deviation of the performance metric. The benchmark model is filled in black, while simple-form candidate models are hollow, and full-form candidate models are filled in gray. The vertical gray bar in the MAE graph is intended to illustrate the approximate size of the Region of Practical Equivalence employed in HDI-ROPE tests. However, it is worth noting the HDI-ROPE test was applied in each region independently (as opposed to using a single ROPE across all regions as shown above). The HDI-ROPE test was also applied using the 90% highest posterior density interval, as opposed to the mean $\pm$ the standard deviation.

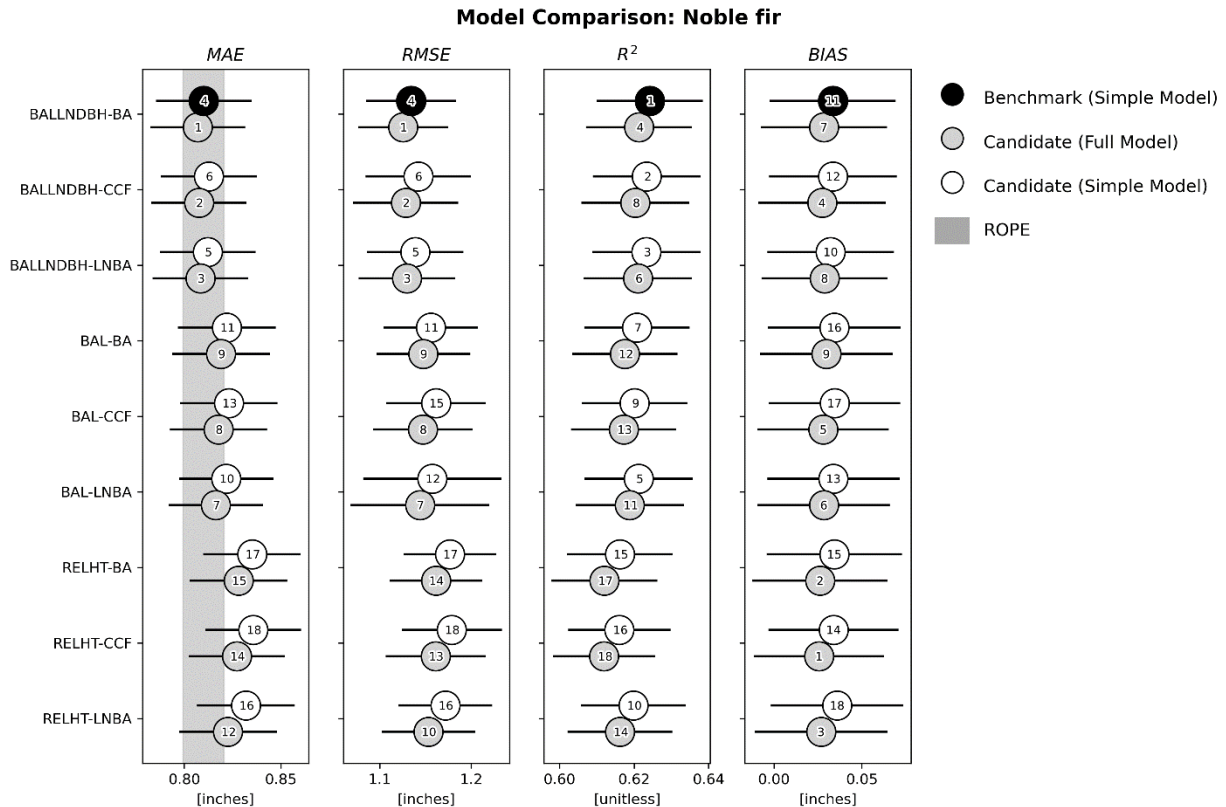


Figure S3.25. Unpooled model comparison with multiple performance metrics: Noble fir. These graphs illustrate the performance of the unpooled benchmark model (SIMPLE-BALLNDBH-BAL) as well as the 17 other unpooled candidate models using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Bayesian R^2 , and average bias. These performance metrics are calculated from all tree observations across the four focal FVS Variants. The combination of stand-level and tree-level competition variables used in each model are labeled at the left side of the graphs. The number inside each dot indicates the rank of the model based on the model's average for that performance metric across the four focal FVS Variants. The location of each dot along the x-axis indicates the mean value for the performance metric, and horizontal lines (if visible beyond the dot) indicate the standard deviation of the performance metric. The benchmark model is filled in black, while simple-form candidate models are hollow, and full-form candidate models are filled in gray. The vertical gray bar in the MAE graph is intended to illustrate the approximate size of the Region of Practical Equivalence employed in HDI-ROPE tests. However, it is worth noting the HDI-ROPE test was applied in each region independently (as opposed to using a single ROPE across all regions as shown above). The HDI-ROPE test was also applied using the 90% highest posterior density interval, as opposed to the mean $\pm$ the standard deviation.

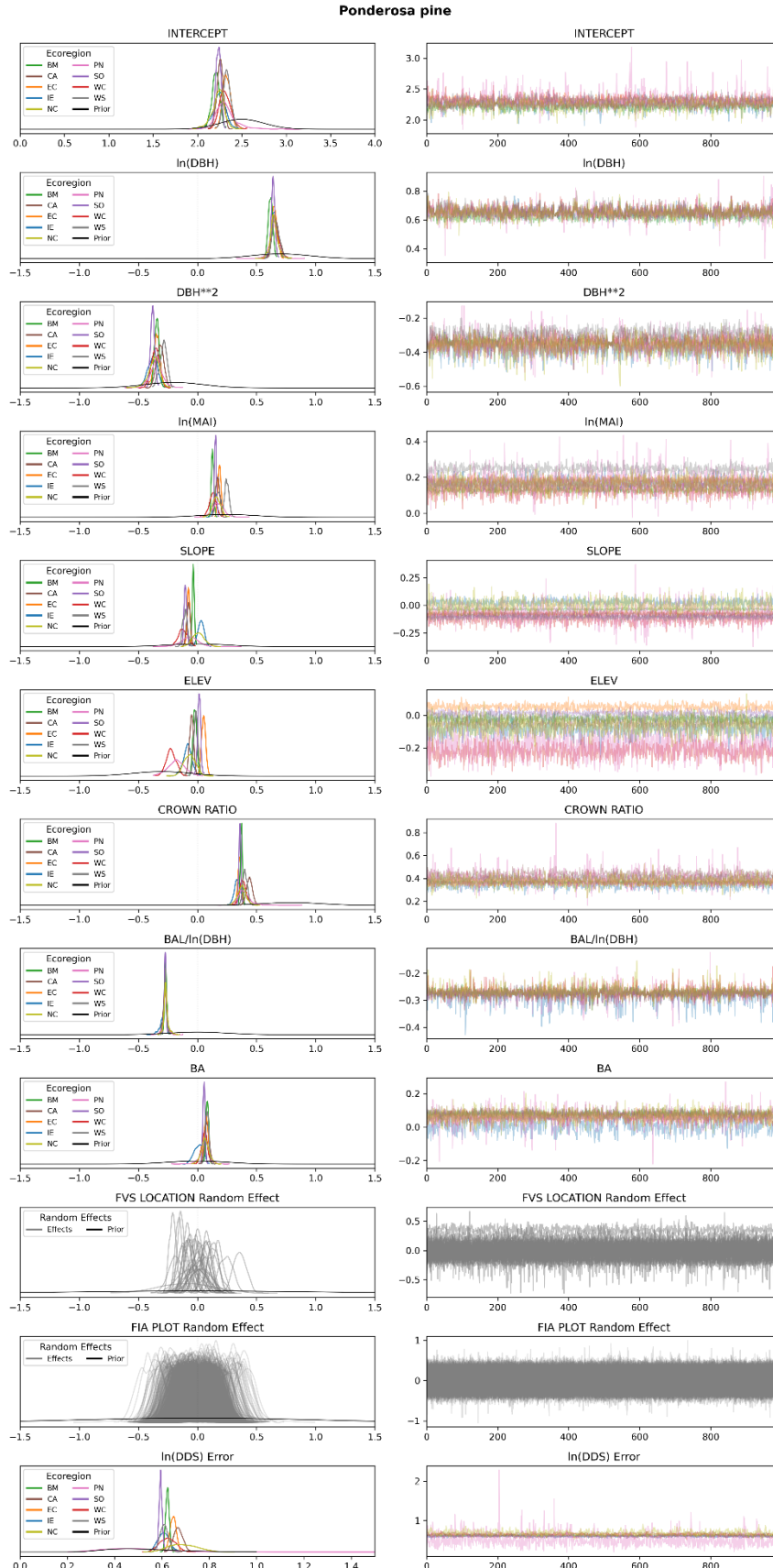


Figure S3.26. Posterior distributions of parameters for the hierarchical benchmark model for Ponderosa pine.

Posterior distributions for each parameter of the partially pooled SIMPLE-BALLNDBH-BA model are displayed in two ways: kernel density plots of the probability density for the parameter values are shown in the left column of plots; and trace plots showing the values from the parameter over 1,000 steps of the MCMC sampler in the right column of plots. Colors are employed in kernel density plots and trace plots to distinguish the parameters fit for each FVS Variant (shown as “Ecoregion” in the figure legends). The prior distribution used for each parameter is shown as a black line in the kernel density plots. The means and standard deviations of each parameter visualized here are reported in Table 3.7. Parameter values displayed are for standardized covariates to allow the sign and magnitude of regression coefficients to be compared.

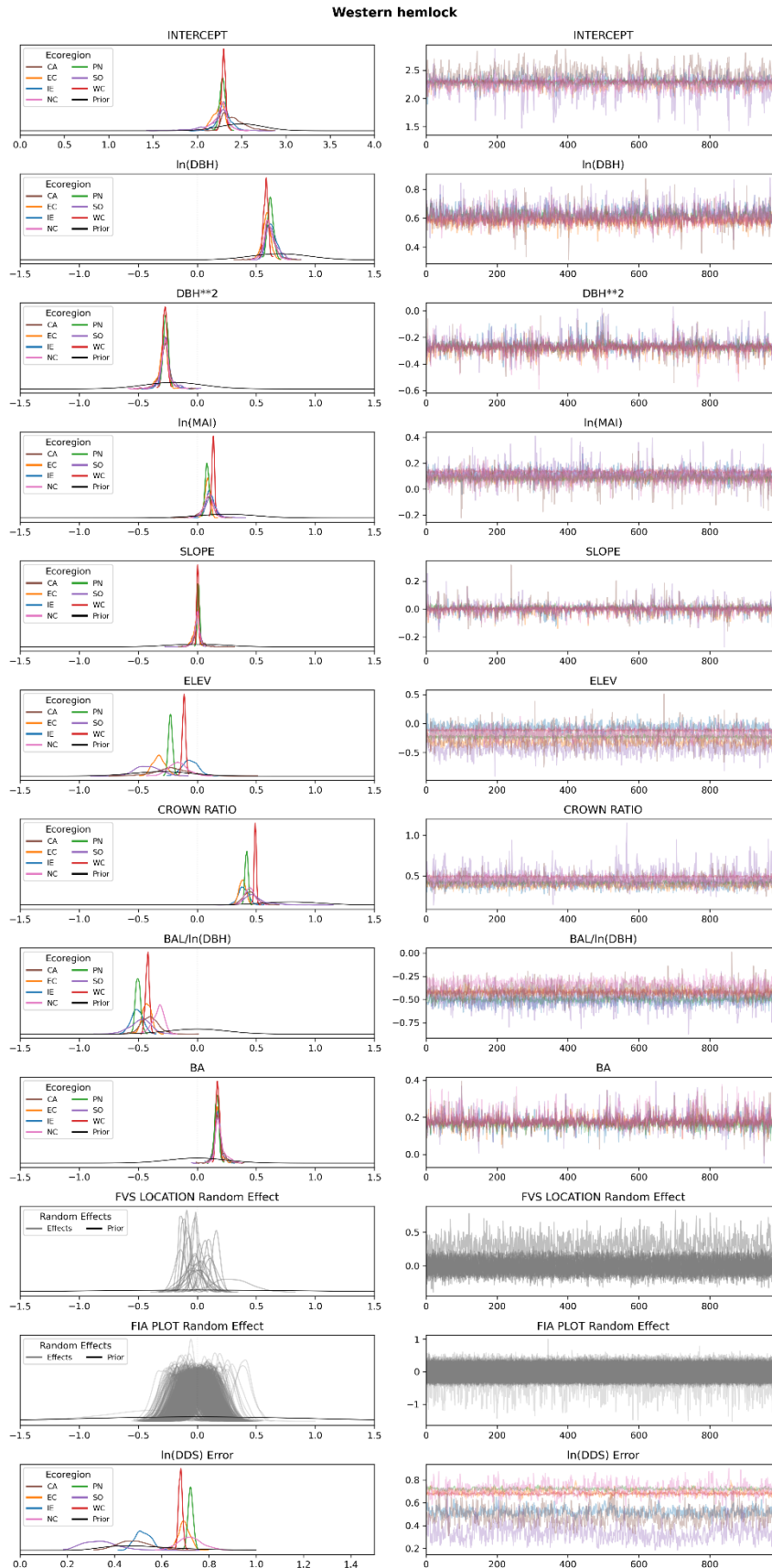


Figure S3.27. Posterior distributions of parameters for the hierarchical benchmark model for Western hemlock.

Posterior distributions for each parameter of the partially pooled SIMPLE-BALLNDBH-BA model are displayed in two ways: kernel density plots of the probability density for the parameter values are shown in the left column of plots; and trace plots showing the values from the parameter over 1,000 steps of the MCMC sampler in the right column of plots. Colors are employed in kernel density plots and trace plots to distinguish the parameters fit for each FVS Variant (shown as “Ecoregion” in the figure legends). The prior distribution used for each parameter is shown as a black line in the kernel density plots. The means and standard deviations of each parameter visualized here are reported in Table 3.7. Parameter values displayed are for standardized covariates to allow the sign and magnitude of regression coefficients to be compared.

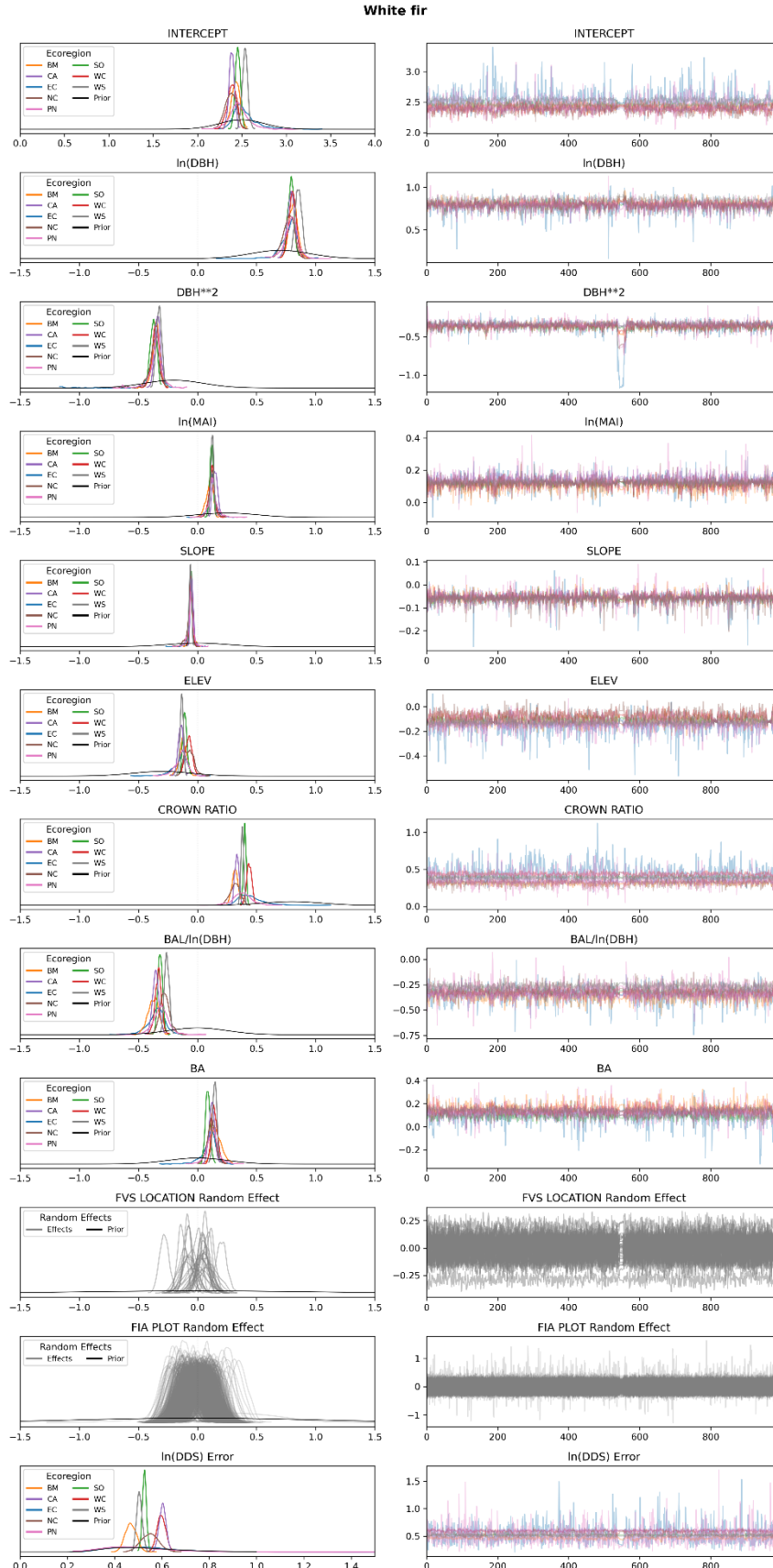


Figure S3.28. Posterior distributions of parameters for the hierarchical benchmark model for White fir.

Posterior distributions for each parameter of the partially pooled SIMPLE-BALLNDBH-BA model are displayed in two ways: kernel density plots of the probability density for the parameter values are shown in the left column of plots; and trace plots showing the values from the parameter over 1,000 steps of the MCMC sampler in the right column of plots. Colors are employed in kernel density plots and trace plots to distinguish the parameters fit for each FVS Variant (shown as “Ecoregion” in the figure legends). The prior distribution used for each parameter is shown as a black line in the kernel density plots. The means and standard deviations of each parameter visualized here are reported in Table 3.7. Parameter values displayed are for standardized covariates to allow the sign and magnitude of regression coefficients to be compared.

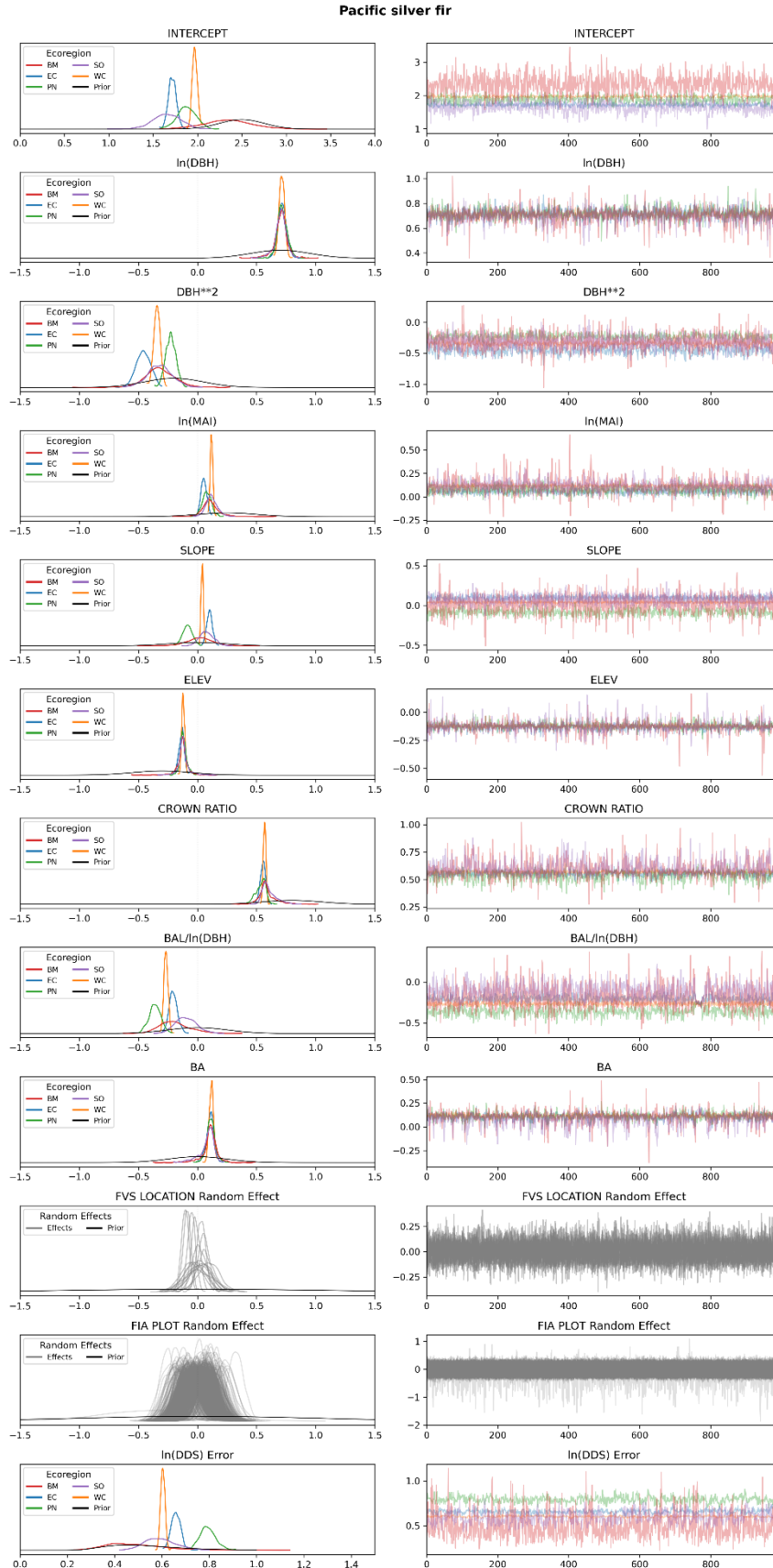


Figure S3.29. Posterior distributions of parameters for the hierarchical benchmark model for Pacific silver fir.

Posterior distributions for each parameter of the partially pooled SIMPLE-BALLNDBH-BA model are displayed in two ways: kernel density plots of the probability density for the parameter values are shown in the left column of plots; and trace plots showing the values from the parameter over 1,000 steps of the MCMC sampler in the right column of plots. Colors are employed in kernel density plots and trace plots to distinguish the parameters fit for each FVS Variant (shown as “Ecoregion” in the figure legends). The prior distribution used for each parameter is shown as a black line in the kernel density plots. The means and standard deviations of each parameter visualized here are reported in Table 3.7. Parameter values displayed are for standardized covariates to allow the sign and magnitude of regression coefficients to be compared.

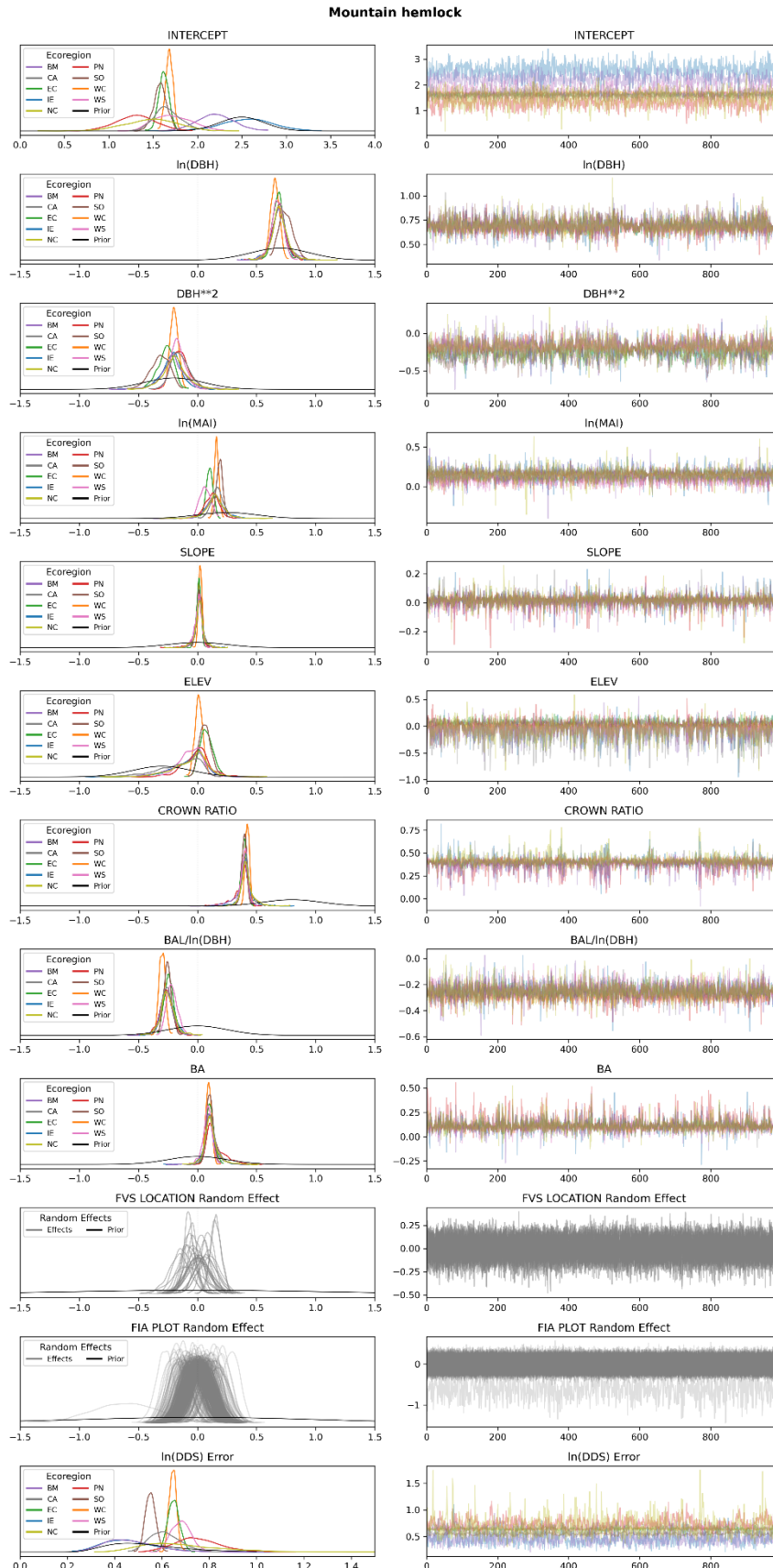


Figure S3.30. Posterior distributions of parameters for the hierarchical benchmark model for Mountain hemlock. Posterior distributions for each parameter of the partially pooled SIMPLE-BALLNDBH-BA model are displayed in two ways: kernel density plots of the probability density for the parameter values are shown in the left column of plots; and trace plots showing the values from the parameter over 1,000 steps of the MCMC sampler in the right column of plots. Colors are employed in kernel density plots and trace plots to distinguish the parameters fit for each FVS Variant (shown as “Ecoregion” in the figure legends). The prior distribution used for each parameter is shown as a black line in the kernel density plots. The means and standard deviations of each parameter visualized here are reported in Table 3.7. Parameter values displayed are for standardized covariates to allow the sign and magnitude of regression coefficients to be compared.

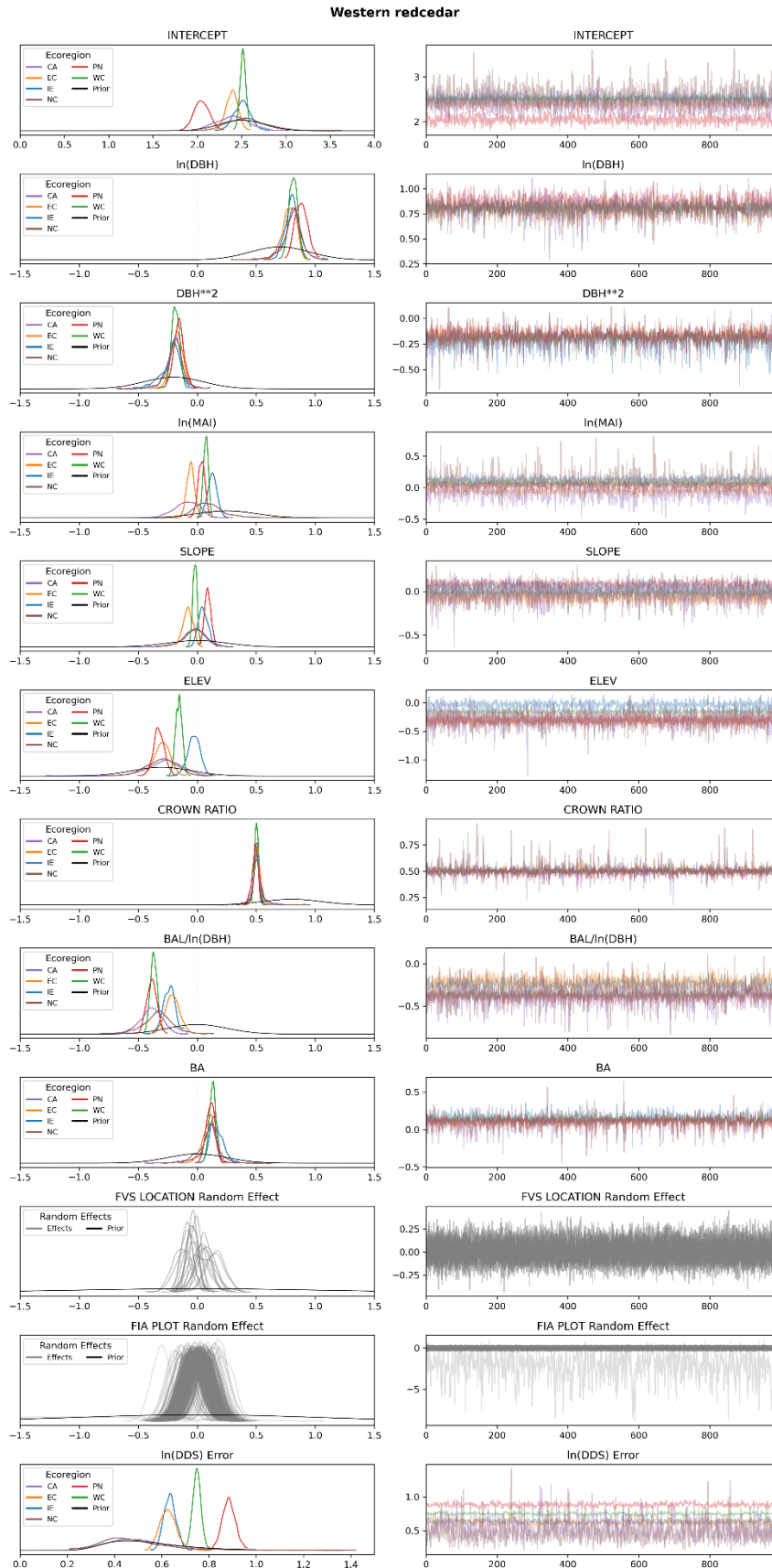


Figure S3.31. Posterior distributions of parameters for the hierarchical benchmark model for Western redcedar.

Posterior distributions for each parameter of the partially pooled SIMPLE-BALLNDBH-BA model are displayed in two ways: kernel density plots of the probability density for the parameter values are shown in the left column of plots; and trace plots showing the values from the parameter over 1,000 steps of the MCMC sampler in the right column of plots. Colors are employed in kernel density plots and trace plots to distinguish the parameters fit for each FVS Variant (shown as “Ecoregion” in the figure legends). The prior distribution used for each parameter is shown as a black line in the kernel density plots. The means and standard deviations of each parameter visualized here are reported in Table 3.7. Parameter values displayed are for standardized covariates to allow the sign and magnitude of regression coefficients to be compared.

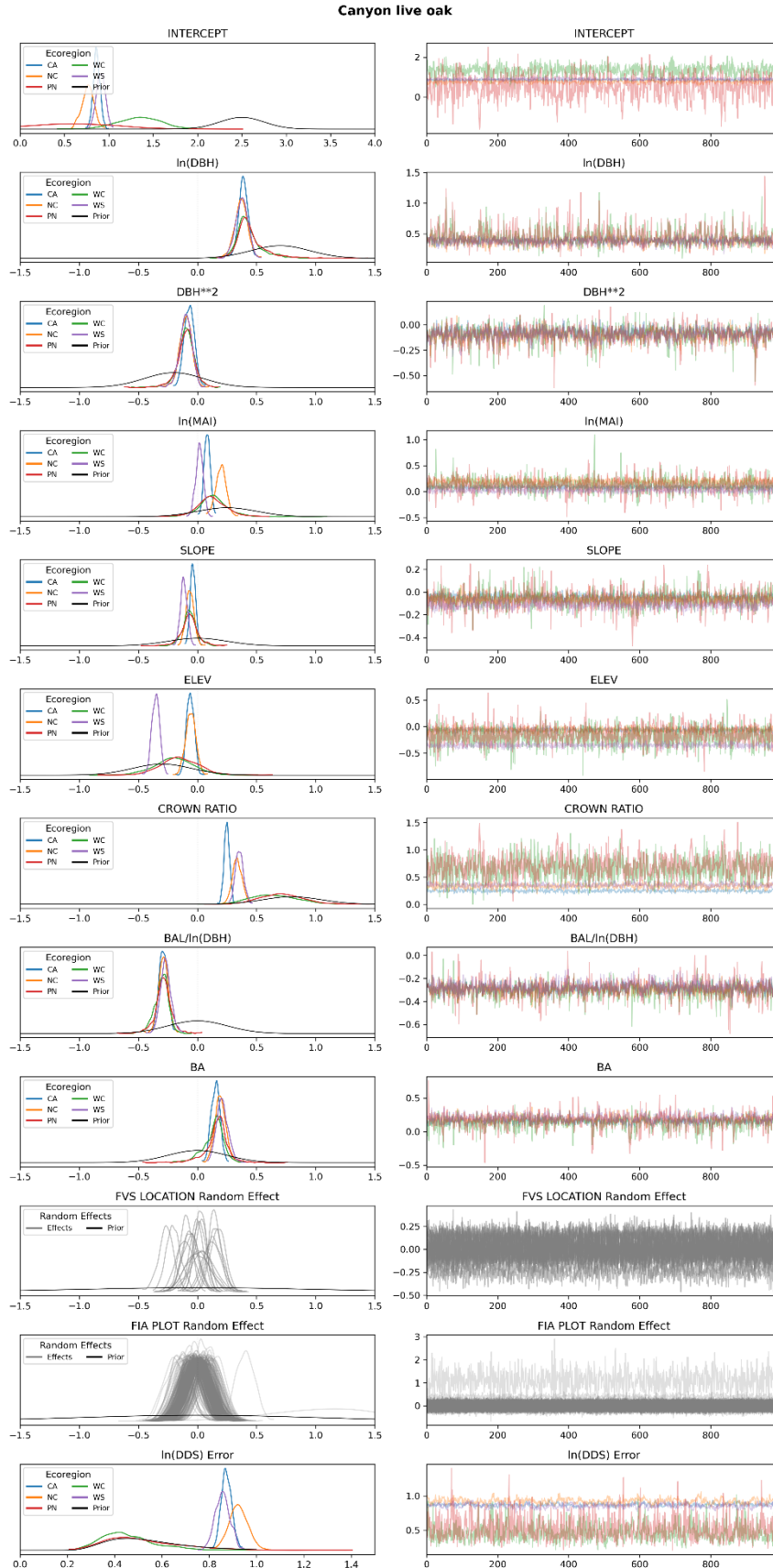


Figure S3.32. Posterior distributions of parameters for the hierarchical benchmark model for Canyon live oak.

Posterior distributions for each parameter of the partially pooled SIMPLE-BALLNDBH-BA model are displayed in two ways: kernel density plots of the probability density for the parameter values are shown in the left column of plots; and trace plots showing the values from the parameter over 1,000 steps of the MCMC sampler in the right column of plots. Colors are employed in kernel density plots and trace plots to distinguish the parameters fit for each FVS Variant (shown as “Ecoregion” in the figure legends). The prior distribution used for each parameter is shown as a black line in the kernel density plots. The means and standard deviations of each parameter visualized here are reported in Table 3.7. Parameter values displayed are for standardized covariates to allow the sign and magnitude of regression coefficients to be compared.

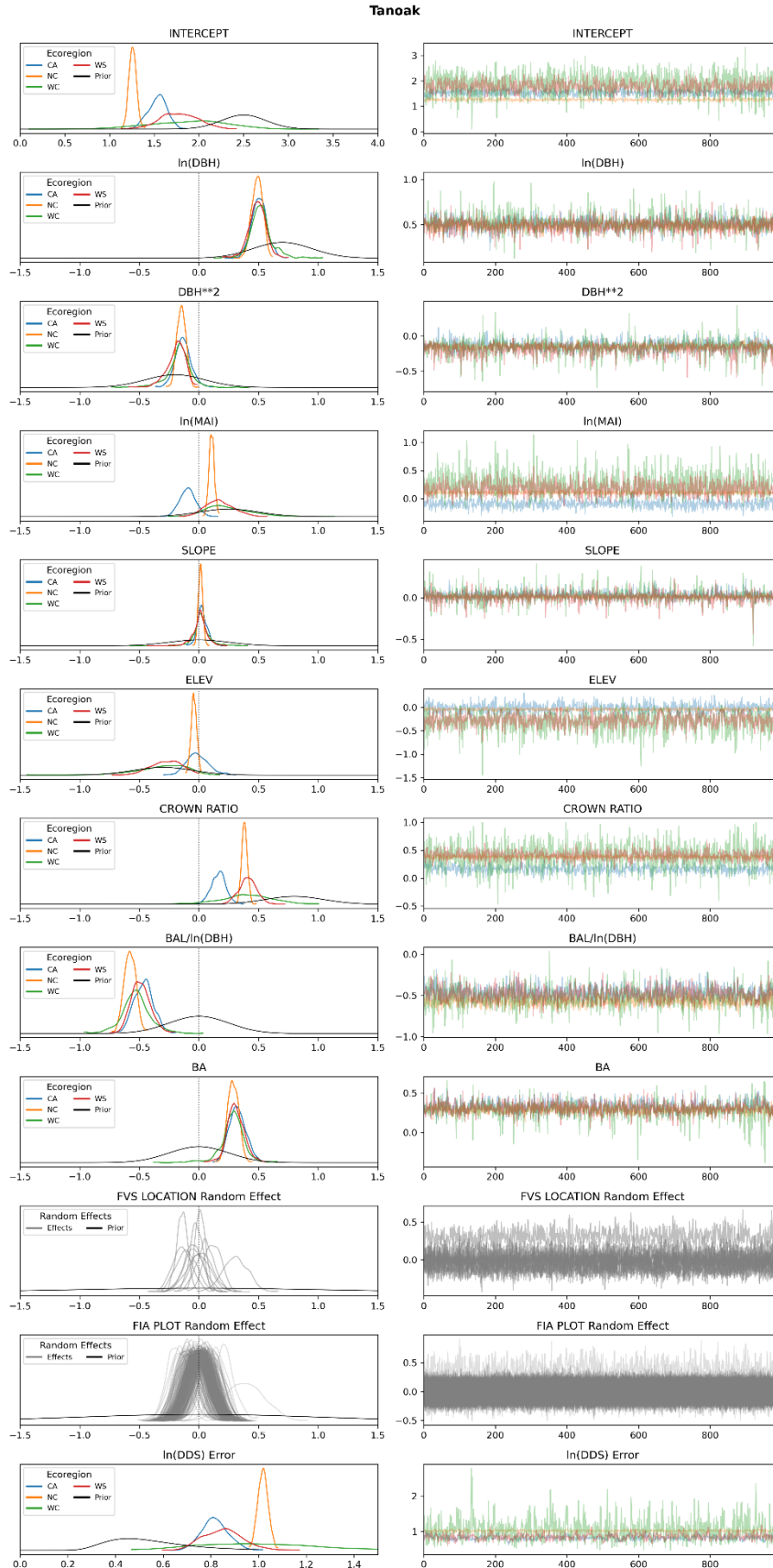


Figure S3.33. Posterior distributions of parameters for the hierarchical benchmark model for Tanoak.

Posterior distributions for each parameter of the partially pooled SIMPLE-BALLNDBH-BA model are displayed in two ways: kernel density plots of the probability density for the parameter values are shown in the left column of plots; and trace plots showing the values from the parameter over 1,000 steps of the MCMC sampler in the right column of plots. Colors are employed in kernel density plots and trace plots to distinguish the parameters fit for each FVS Variant (shown as “Ecoregion” in the figure legends). The prior distribution used for each parameter is shown as a black line in the kernel density plots. The means and standard deviations of each parameter visualized here are reported in Table 3.7. Parameter values displayed are for standardized covariates to allow the sign and magnitude of regression coefficients to be compared.

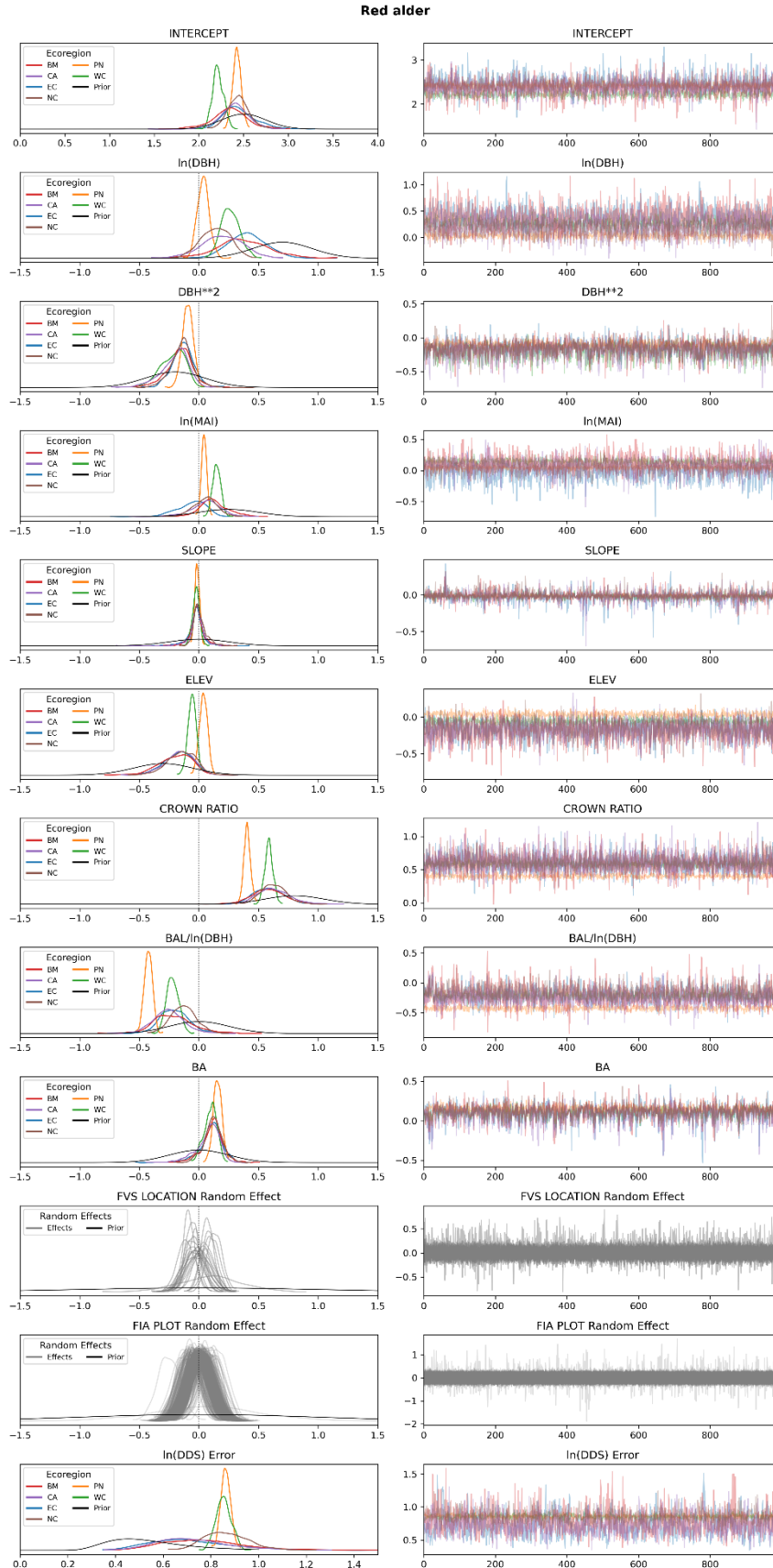


Figure S3.34. Posterior distributions of parameters for the hierarchical benchmark model for Red alder.

Posterior distributions for each parameter of the partially pooled SIMPLE-BALLNDBH-BA model are displayed in two ways: kernel density plots of the probability density for the parameter values are shown in the left column of plots; and trace plots showing the values from the parameter over 1,000 steps of the MCMC sampler in the right column of plots. Colors are employed in kernel density plots and trace plots to distinguish the parameters fit for each FVS Variant (shown as “Ecoregion” in the figure legends). The prior distribution used for each parameter is shown as a black line in the kernel density plots. The means and standard deviations of each parameter visualized here are reported in Table 3.7. Parameter values displayed are for standardized covariates to allow the sign and magnitude of regression coefficients to be compared.

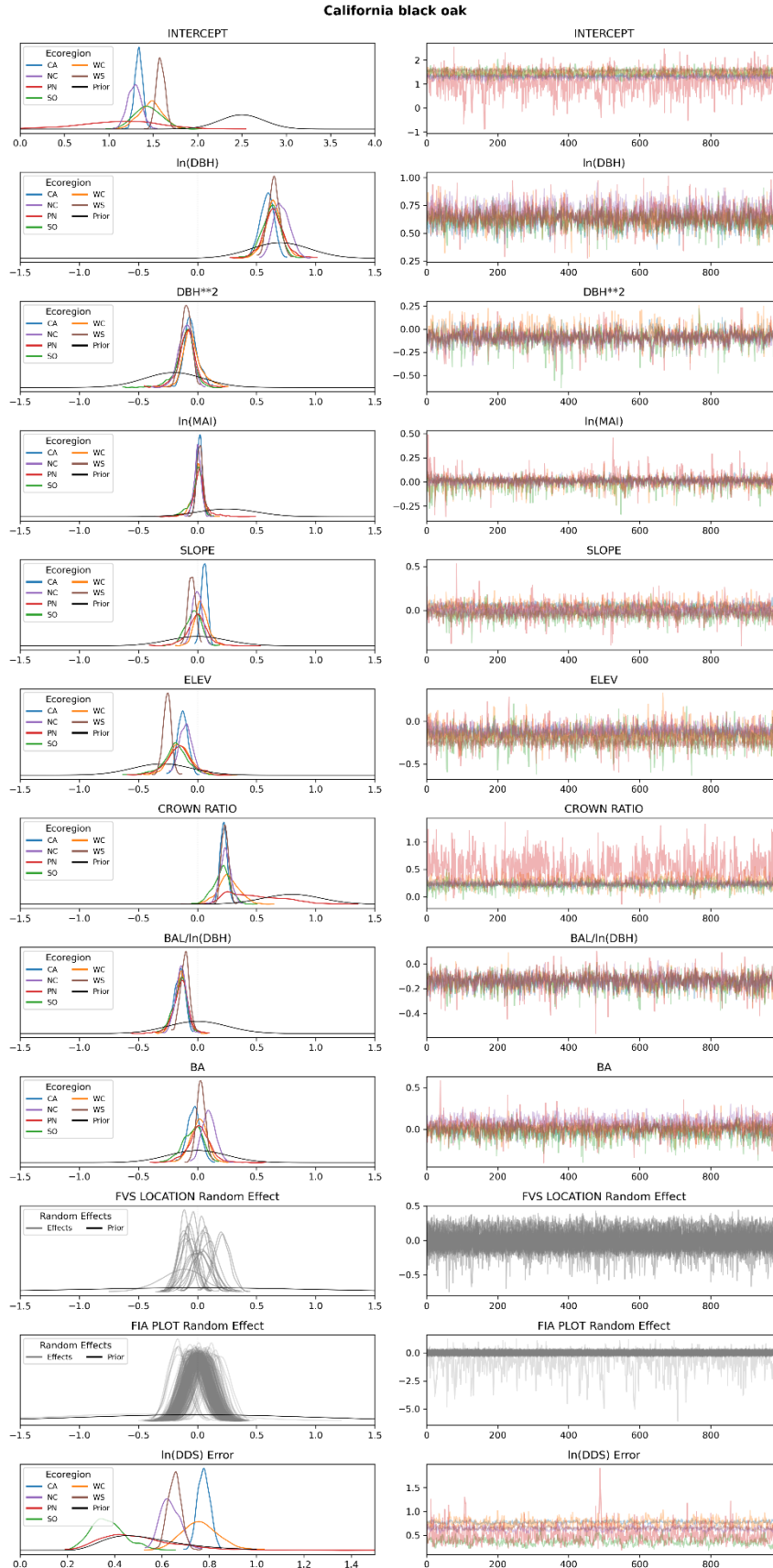


Figure S3.35. Posterior distributions of parameters for the hierarchical benchmark model for California black oak. Posterior distributions for each parameter of the partially pooled SIMPLE-BALLNDBH-BA model are displayed in two ways: kernel density plots of the probability density for the parameter values are shown in the left column of plots; and trace plots showing the values from the parameter over 1,000 steps of the MCMC sampler in the right column of plots. Colors are employed in kernel density plots and trace plots to distinguish the parameters fit for each FVS Variant (shown as “Ecoregion” in the figure legends). The prior distribution used for each parameter is shown as a black line in the kernel density plots. The means and standard deviations of each parameter visualized here are reported in Table 3.7. Parameter values displayed are for standardized covariates to allow the sign and magnitude of regression coefficients to be compared.

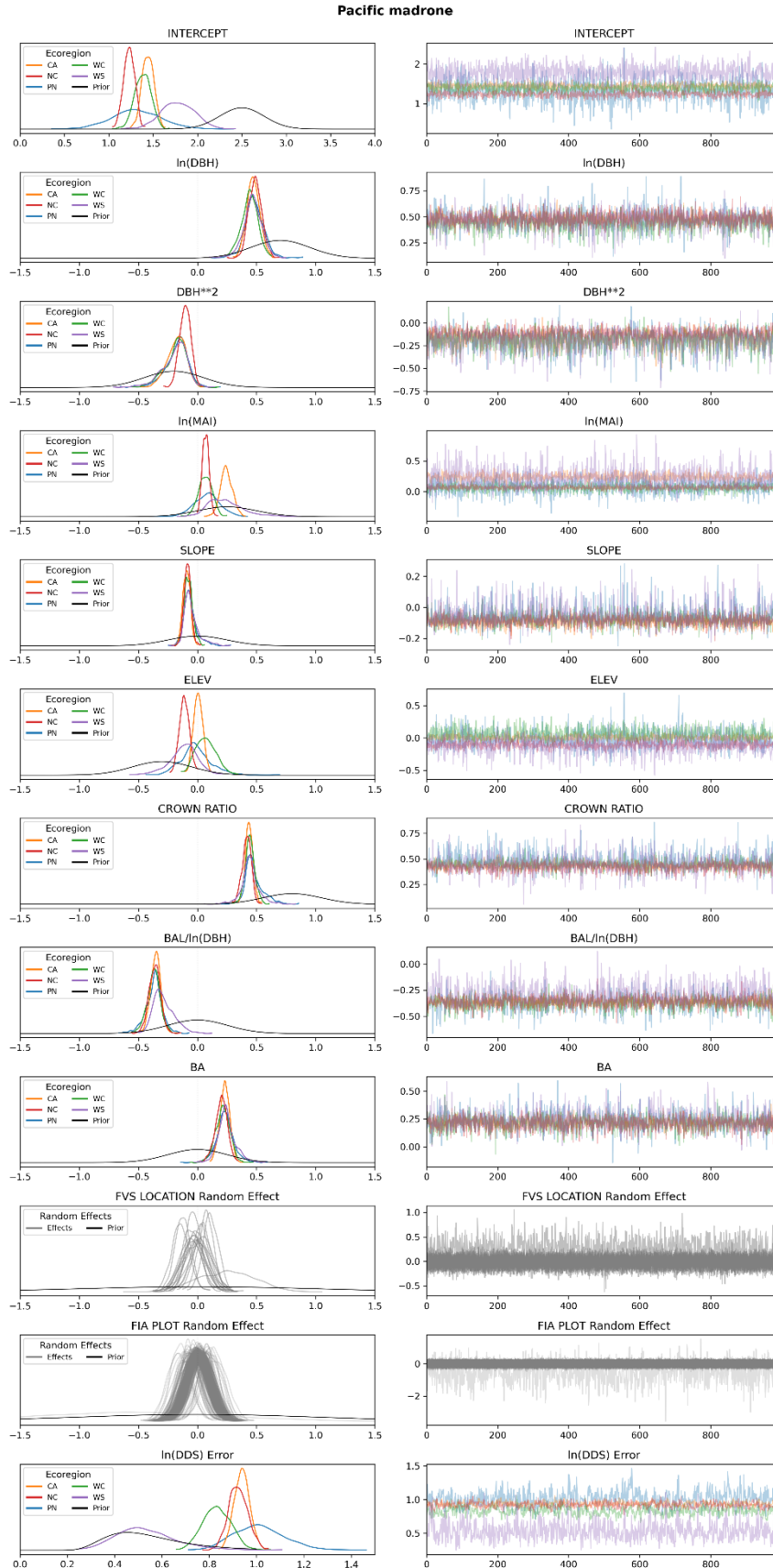


Figure S3.36. Posterior distributions of parameters for the hierarchical benchmark model for Pacific madrone.

Posterior distributions for each parameter of the partially pooled SIMPLE-BALLNDBH-BA model are displayed in two ways: kernel density plots of the probability density for the parameter values are shown in the left column of plots; and trace plots showing the values from the parameter over 1,000 steps of the MCMC sampler in the right column of plots. Colors are employed in kernel density plots and trace plots to distinguish the parameters fit for each FVS Variant (shown as “Ecoregion” in the figure legends). The prior distribution used for each parameter is shown as a black line in the kernel density plots. The means and standard deviations of each parameter visualized here are reported in Table 3.7. Parameter values displayed are for standardized covariates to allow the sign and magnitude of regression coefficients to be compared.

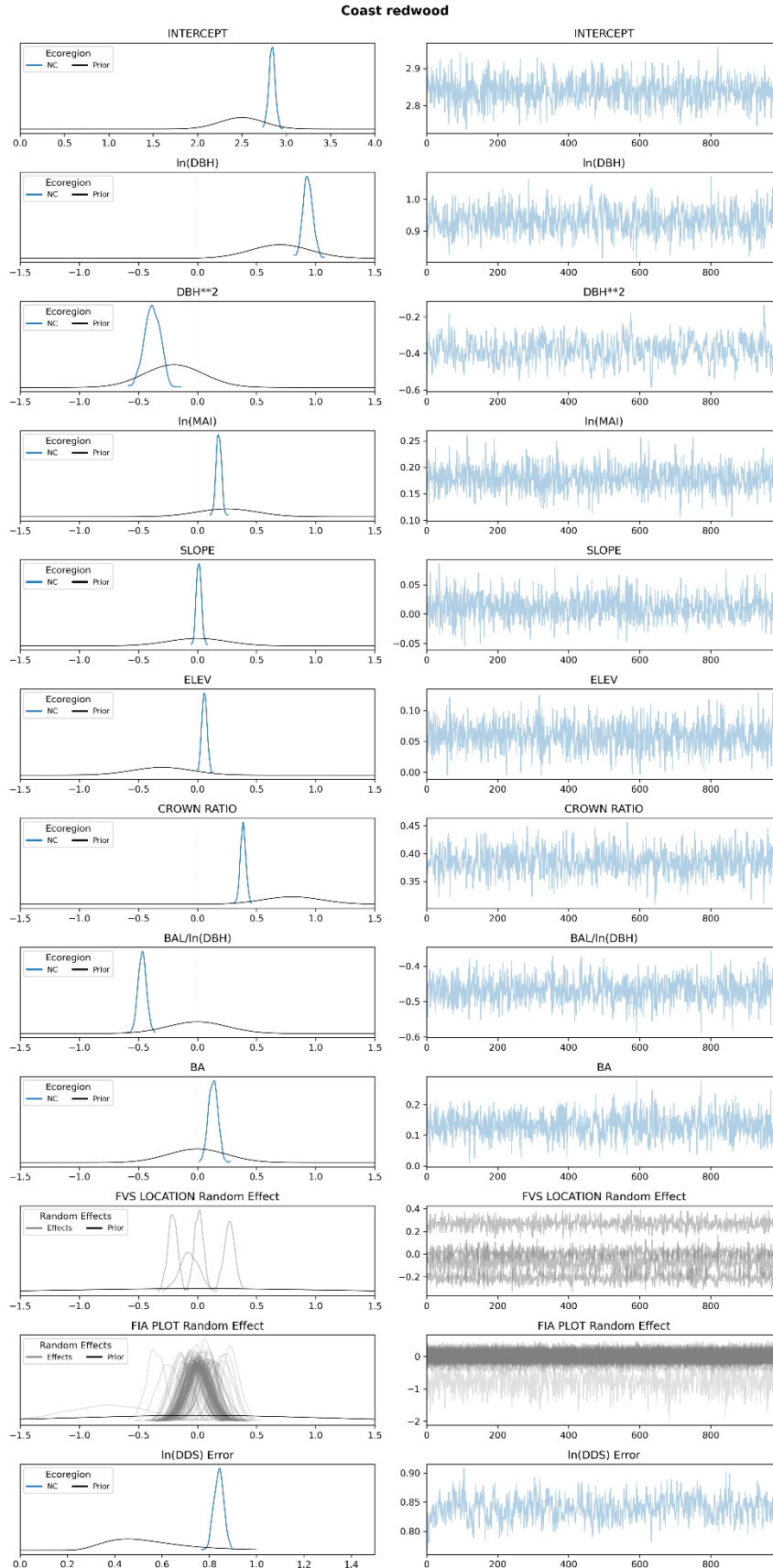


Figure S3.37. Posterior distributions of parameters for the hierarchical benchmark model for Coast redwood.

Posterior distributions for each parameter of the partially pooled SIMPLE-BALLNDBH-BA model are displayed in two ways: kernel density plots of the probability density for the parameter values are shown in the left column of plots; and trace plots showing the values from the parameter over 1,000 steps of the MCMC sampler in the right column of plots. Colors are employed in kernel density plots and trace plots to distinguish the parameters fit for each FVS Variant (shown as “Ecoregion” in the figure legends). The prior distribution used for each parameter is shown as a black line in the kernel density plots. The means and standard deviations of each parameter visualized here are reported in Table 3.7. Parameter values displayed are for standardized covariates to allow the sign and magnitude of regression coefficients to be compared.

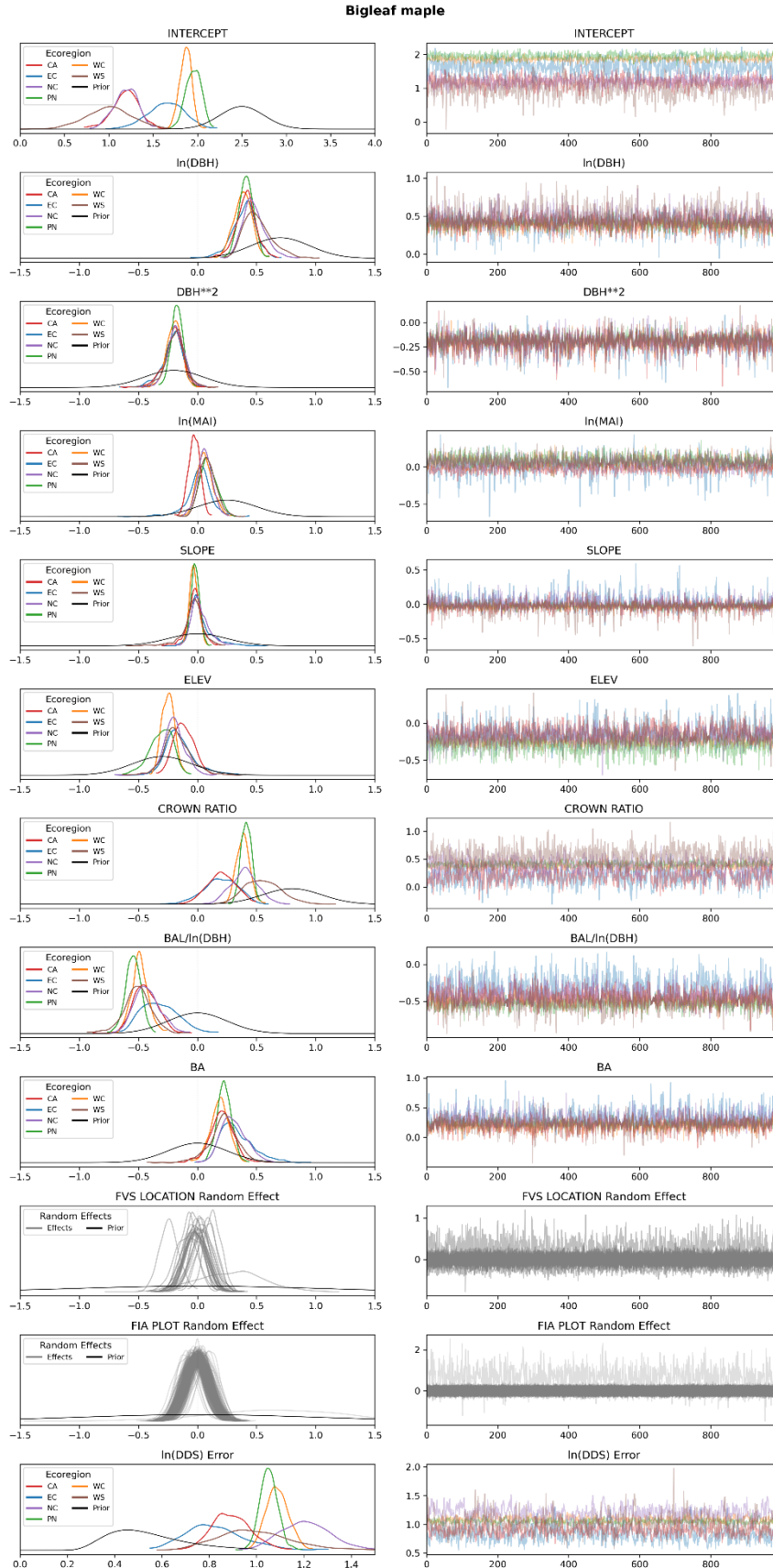


Figure S3.38. Posterior distributions of parameters for the hierarchical benchmark model for Bigleaf maple.

Posterior distributions for each parameter of the partially pooled SIMPLE-BALLNDBH-BA model are displayed in two ways: kernel density plots of the probability density for the parameter values are shown in the left column of plots; and trace plots showing the values from the parameter over 1,000 steps of the MCMC sampler in the right column of plots. Colors are employed in kernel density plots and trace plots to distinguish the parameters fit for each FVS Variant (shown as “Ecoregion” in the figure legends). The prior distribution used for each parameter is shown as a black line in the kernel density plots. The means and standard deviations of each parameter visualized here are reported in Table 3.7. Parameter values displayed are for standardized covariates to allow the sign and magnitude of regression coefficients to be compared.

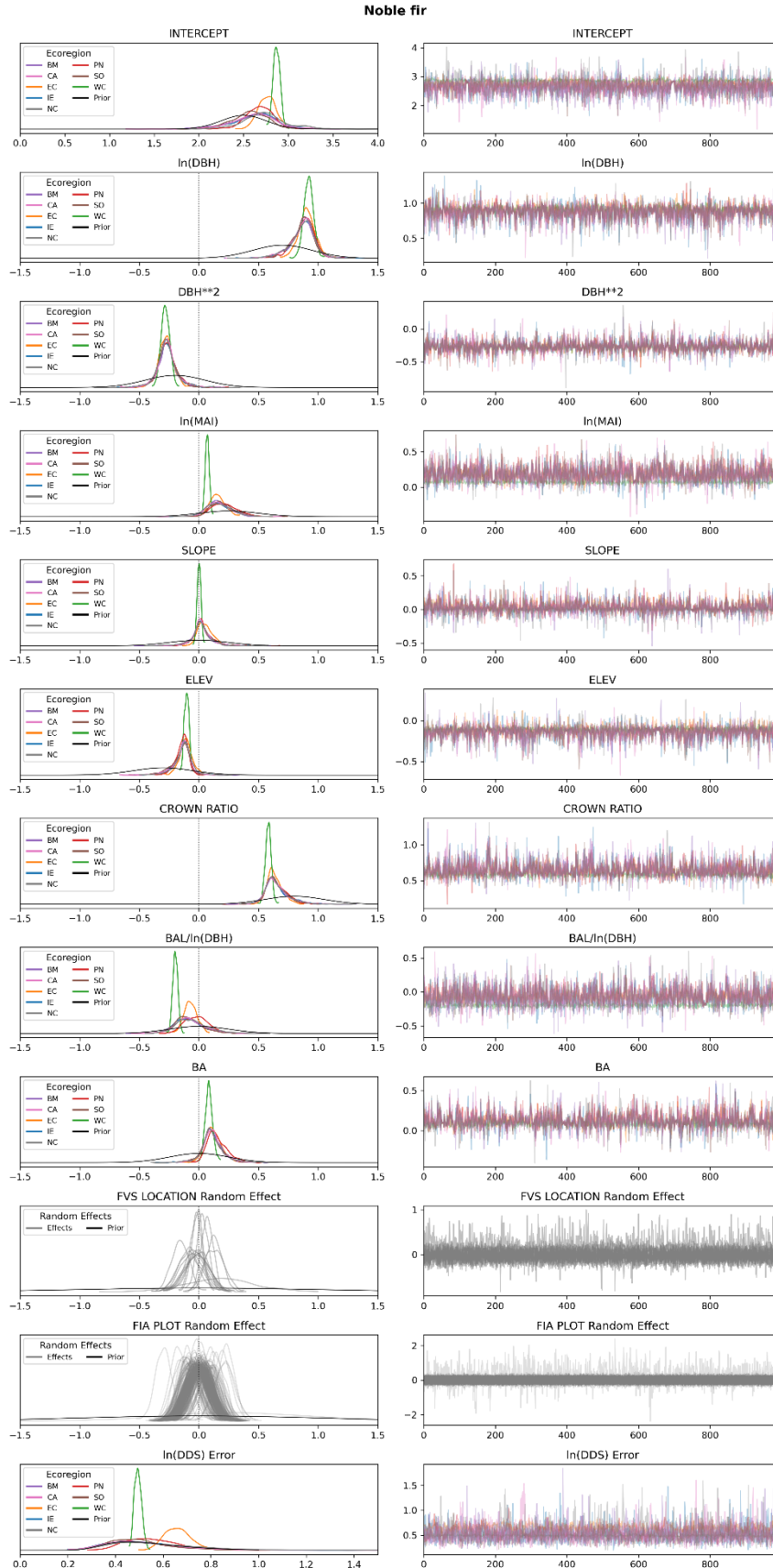


Figure S3.39. Posterior distributions of parameters for the hierarchical benchmark model for Noble fir.

Posterior distributions for each parameter of the partially pooled SIMPLE-BALLNDBH-BA model are displayed in two ways: kernel density plots of the probability density for the parameter values are shown in the left column of plots; and trace plots showing the values from the parameter over 1,000 steps of the MCMC sampler in the right column of plots. Colors are employed in kernel density plots and trace plots to distinguish the parameters fit for each FVS Variant (shown as “Ecoregion” in the figure legends). The prior distribution used for each parameter is shown as a black line in the kernel density plots. The means and standard deviations of each parameter visualized here are reported in Table 3.7. Parameter values displayed are for standardized covariates to allow the sign and magnitude of regression coefficients to be compared.

REFERENCES CITED

- Adams, H., Williams, A., Xu, C., Rauscher, S., Jiang, X., McDowell, N., 2013. Empirical and process-based approaches to climate-induced forest mortality models. *Front. Plant Sci.* 4.
- Agne, M.C., Beedlow, P.A., Shaw, D.C., Woodruff, D.R., Lee, E.H., Cline, S.P., Comeleo, R.L., 2018. Interactions of predominant insects and diseases with climate change in Douglas-fir forests of western Oregon and Washington, U.S.A. *For. Ecol. Manag.* 409, 317–332. <https://doi.org/10.1016/j.foreco.2017.11.004>
- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control* 19, 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- American Carbon Registry, 2018. The American Carbon Registry Standard (No. version 5.0). Winrock International, Arlington, VA.
- Anderson, P., Battaglia, M., Johnson, M., Kabrick, J., Shaw, J., 2020. Research Applications of the Forest Vegetation Simulator: Documentation and Citation. USDA Forest Service, Forest Management Service Center, Fort Collins, CO.
- Arciniega, J., 2017. Using Landfire, FSVeg Spatial Data Analyzer Nearest Neighbor, Forest Vegetation Simulator, and FlamMap to (General Technical Report No. SRS-GTR-224), Proceedings of the 2017 Forest Vegetation Simulator (FVS) e-Conference. U.S. Department of Agriculture Forest Service, Southern Research Station, Asheville, NC.
- Arney, J.D., 2016. The Economic Results of a PNW Silvicultural Costs Survey: Are You Swimming Above or Below the Financial Waterline?, in: 2016 PNW Reforestation Council Annual Meeting (Vancouver, WA). Forest Biometrics Research Institute, Portland, OR, p. 41.
- Atkins, D., Lundberg, R., 2002. Analyst Hazards When Assessing Fire, Insect, and Disease Hazard in Montana Using FIA Data with FVS: Or Alligators We Didn't See Coming (Proceedings of the Second Forest Vegetation Simulator Conference, February 2002 No. RMRS-P-25). U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Ogden, UT.
- Bagdon, B., Huang, C.-H., 2014. Carbon Stocks and Climate Change: Management Implications in Northern Arizona Ponderosa Pine Forests. *Forests* 5, 620–642. <https://doi.org/10.3390/f5040620>
- Bagdon, B.A., Nguyen, T.H., Vorster, A., Paustian, K., Field, J.L., 2021. A model evaluation framework applied to the Forest Vegetation Simulator (FVS) in Colorado and Wyoming lodgepole pine forests. *For. Ecol. Manag.* 480, 118619. <https://doi.org/10.1016/j.foreco.2020.118619>
- Bergman, R., Taylor, A., 2011. EPD-Environmental Product Declarations for Wood Products—an Application of Life Cycle Information About Forest Products. *For. Prod. J.* 61, 192–201. <https://doi.org/10.13073/0015-7473-61.3.192>
- Bigsby, H., 2009. Carbon banking: Creating flexibility for forest owners. *For. Ecol. Manag.* 257, 378–383. <https://doi.org/10.1016/j.foreco.2008.09.018>

- Binkley, C.S., 2007. The rise and fall of the timber investment management organizations: Ownership changes in US forestlands, in: 2007 Pinchot Distinguished Lecture. Pinchot Institute for Conservation, Washington, D.C., p. 12.
- Blacklock, N., 2016. 2016 Strategic Issues for US Pacific Northwest Timberlands, in: 3rd Annual Western Forest Industry Conference: Mapping the Course – Timberlands, Forest Products Processing and Energy Issues for 2016. Western Forestry and Conservation Association, Vancouver, WA, p. 23.
- Bliss, J.C., Kelly, E.C., Abrams, J., Bailey, C., Dyer, J., 2010. Disintegration of the U. S. Industrial Forest Estate: Dynamics, Trajectories, and Questions. *Small-Scale For.* 9, 53–66. <https://doi.org/10.1007/s11842-009-9101-7>
- Bouslah, K., M’Zali, B., Turcotte, M.-F., Kooli, M., 2010. The Impact of Forest Certification on Firm Financial Performance in Canada and the U.S. *J. Bus. Ethics* 96, 551–572. <https://doi.org/10.1007/s10551-010-0482-5>
- Bradbury, J., Frostig, R., Hawkins, P., Johnson, M.J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., Zhang, Q., 2018. JAX: composable transformations of Python+NumPy programs.
- Brazee, R.J., 2001. Introduction—The Faustmann Formula: Fundamental to Forest Economics 150 Years After Publication. *For. Sci.* 47, 441–442. <https://doi.org/10.1093/forestscience/47.4.441>
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Bruce, D., 1977. Yield Differences between Research Plots and Managed Forests. *J. For.* 75, 14–17. <https://doi.org/10.1093/jof/75.1.14>
- Cacho, O.J., Lipper, L., Moss, J., 2013. Transaction costs of carbon offset projects: A comparative study. *Ecol. Econ., Transaction Costs and Environmental Policy* 88, 232–243. <https://doi.org/10.1016/j.ecolecon.2012.12.008>
- Cacho, O.J., Wise, R.M., MacDicken, K.G., 2004. Carbon Monitoring Costs and their Effect on Incentives to Sequester Carbon through Forestry. *Mitig. Adapt. Strateg. Glob. Change* 9, 273–293. <https://doi.org/10.1023/B:MITI.0000029930.11262.b8>
- Cailleret, M., Bircher, N., Hartig, F., Hülsmann, L., Bugmann, H., 2020. Bayesian calibration of a growth-dependent tree mortality model to simulate the dynamics of European temperate forests. *Ecol. Appl.* 30, 1–17.
- California Air Resources Board, 2015. Compliance Offset Protocol: U.S. Forest Projects. California Environmental Protection Agency, Air Resources Board, Sacramento, CA.
- Canavan, S.J., Ramm, C.W., 2000. Accuracy and Precision of 10 Year Predictions for Forest Vegetation Simulator—Lake States. *North. J. Appl. For.* 17, 62–70. <https://doi.org/10.1093/njaf/17.2.62>
- Cao, Q.V., 2004. Annual tree growth predictions from periodic measurements, in: Proceedings of the 12th Biennial Southern Silvicultural Research Conference. Gen. Tech. Rep. SRS–71. USDA Forest Service Southern Research Station, Asheville, NC, Biloxi, Mississippi, pp. 212–215. <https://doi.org/10.1093/forestscience/46.1.127>
- Cao, Q.V., 2000. Prediction of Annual Diameter Growth and Survival for Individual Trees from Periodic Measurements. *For. Sci.* 46, 127–131. <https://doi.org/10.1093/forestscience/46.1.127>

- Cash, D.W., Clark, W.C., Alcock, F., Dickson, N.M., Eckley, N., Guston, D.H., Jäger, J., Mitchell, R.B., 2003. Knowledge systems for sustainable development. *Proc. Natl. Acad. Sci.* 100, 8086–8091. <https://doi.org/10.1073/pnas.1231332100>
- Cawrse, D., Keyser, C., Keyser, T., Sanchez Meador, A., Smith-Mateja, E., Van Dyck, M., 2010. Forest Vegetation Simulator Model Validation Protocols. US Department of Agriculture, Forest Service, Forest Management Service Center, FVS Steering Team, Validation Subcommittee, Fort Collins, CO.
- CDM - Executive Board, 2007. Tool for the demonstration and assessment of additionality in A/R CDM project activities. (Version 02). (A/R Methodological Tool No. EB 35 annex 17). Clean Development Mechanism Executive Board.
- Chambers, C.J., 1980. Empirical Growth and Yield Tables for the Douglas Fir Zone (No. DNR Report No. 41). Washington Department of Natural Resources, Olympia, WA.
- Charnley, S., Diaz, D., Gosnell, H., 2010. Mitigating Climate Change Through Small-Scale Forestry in the USA: Opportunities and Challenges. *Small-Scale For.* 9, 445–462. <https://doi.org/10.1007/s11842-010-9135-x>
- Clark, J.S., Wolosin, M., Dietze, M., Ibanez, I., LaDeau, S., Welsh, M., Kloeppel, B., 2007. Tree growth inference and prediction from diameter censuses and ring widths. *Ecol. Appl.* 17, 1942–1953.
- Conda contributors, 2022. Conda.
- Cosenza, D.N., Korhonen, L., Maltamo, M., Packalen, P., Strunk, J.L., Næsset, E., Gobakken, T., Soares, P., Tomé, M., 2020. Comparison of linear regression, k-nearest neighbour and random forest methods in airborne laser-scanning-based prediction of growing stock. *For. Int. J. For. Res.* <https://doi.org/10.1093/forestry/cpaa034>
- Cowan, S., Davies, B., Diaz, D., Enelow, N., Halsey, K., Langstaff, K., 2014. Optimizing Urban Ecosystem Services: The Bullitt Center Case Study. Ecotrust, Portland, OR.
- Crookston, N.L., Dixon, G.E., 2005. The forest vegetation simulator: A review of its structure, content, and applications. *Comput. Electron. Agric.* 49, 60–80. <https://doi.org/10.1016/j.compag.2005.02.003>
- Crookston, N.L., Gammel, D.L., Rebain, S., Robinson, D.C.E., Keyser, C., Dahl, C., 2003. Users Guide to the Database Extension of the Forest Vegetation Simulator Version 2.0 (Internal Report). U.S. Department of Agriculture, Forest Service, Forest Management Service Center, Fort Collins, CO.
- Curtis, R.O., 2006. Volume growth trends in a Douglas-fir levels-of-growing-stock study. *West. J. Appl. For.* 212 79-86.
- Curtis, R.O., 1997. The Role of Extended Rotations, in: *Creating a Forestry for the 21st Century: The Science Of Ecosystem Management*. Island Press, Washington, D.C., pp. 165–170.
- Curtis, R.O., 1995. Extended Rotations and Culmination Age of Coast Douglas-fir: Old Studies Speak to Current Issues (Research Paper No. PNW-RP-485). U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR.
- Curtis, R.O., Clendenen, G.W., Reukema, D.L., DeMars, D.J., 1982. Yield Tables for Managed Stands of Coast Douglas-Fir (General Technical Report No. PNW-GTR-135). USDA Forest Service, Pacific Northwest Forest and Range Experiment Station, Portland, OR.
- Curtis, R.O., Marshall, D.D., 2002. Levels-of-growing-stock cooperative study in Douglas-fir: report no. 14—Stampede Creek, 30-year results. (Research Paper No. PNW-RP-543). U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR.

- Curtis, R.O., Marshall, D.D., 1993. Douglas-fir rotations - Time for reappraisal? *West. J. Appl. For.* 8, 81–85.
- Curtis, R.O., Marshall, D.D., Dean, S., others, 2004. Silvicultural options for young-growth Douglas-fir forests: the Capitol Forest study—establishment and first results.
- Cutler, M.E.J., Boyd, D.S., Foody, G.M., Vetrivel, A., 2012. Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: An assessment of predictions between regions. *ISPRS J. Photogramm. Remote Sens.* 70, 66–77. <https://doi.org/10.1016/j.isprsjprs.2012.03.011>
- de Lera Garrido, A., Gobakken, T., Ørka, H., Næsset, E., Bollandsås, O., 2020. Reuse of field data in ALS-assisted forest inventory. *Silva Fenn.* 54. <https://doi.org/10.14214/sf.10272>
- Demšar, J., 2006. Statistical Comparisons of Classifiers over Multiple Data Sets. *J. Mach. Learn. Res.* 7, 1–30. <https://doi.org/10.5555/1248547.1248548>
- Diaz, D., Christoffersen, N., Ettl, G., Hanson, K., 2020. Model-Assisted Forest Stand Delineation to Make Forest Assessment, Valuation, and Management Planning More Accessible (Project Overview No. SW20-914). SARE Grant Management System, Missoula, MT.
- Diaz, D., Perry, M., Tutak, J., Hodges, R., Mertens, M., 2015. Potential climate change impacts on management outcomes for western Oregon BLM forestlands simulated using Climate-FVS. Ecotrust, Portland, OR.
- Diaz, D.D., Loreno, S., Ettl, G.J., Davies, B., 2018. Tradeoffs in Timber, Carbon, and Cash Flow under Alternative Management Systems for Douglas-Fir in the Pacific Northwest. *Forests* 9, 447. <https://doi.org/10.3390/f9080447>
- Dixon, G.E., 2020. Essential FVS: A user’s guide to the Forest Vegetation Simulator (Internal Report). USDA Forest Service, Forest Management Service Center, Fort Collins, CO.
- Dixon, G.E., 2017. Essential FVS: A user’s guide to the Forest Vegetation Simulator (Internal Report). USDA Forest Service, Forest Management Service Center, Fort Collins, CO.
- Ecotrust, 2018. FSC Case Studies [WWW Document]. GitHub Repos. URL https://github.com/Ecotrust/FSC_Case_Studies
- Eriksson, L.O., Sallnäs, O., Ståhl, G., 2007. Forest certification and Swedish wood supply. *For. Policy Econ.* 9, 452–463. <https://doi.org/10.1016/j.forpol.2005.11.001>
- Ewijk, K. van, Tompalski, P., Treitz, P., Coops, N.C., Woods (ret.), M., Pitt (ret.), D., 2020. Transferability of ALS-Derived Forest Resource Inventory Attributes Between an Eastern and Western Canadian Boreal Forest Mixedwood Site. *Can. J. Remote Sens.* 46, 214–236. <https://doi.org/10.1080/07038992.2020.1769470>
- Fekety, P.A., Falkowski, M.J., Hudak, A.T., 2015. Temporal transferability of LiDAR-based imputation of forest inventory attributes. *Can. J. For. Res.* 45, 422–435. <https://doi.org/10.1139/cjfr-2014-0405>
- Fekety, P.A., Falkowski, M.J., Hudak, A.T., Jain, T.B., Evans, J.S., 2018. Transferability of lidar-derived basal area and stem density models within a northern Idaho ecoregion. *Can. J. Remote Sens.* 44, 131–143. <https://doi.org/10.1080/07038992.2018.1461557>
- Finney, M.A., Seli, R.C., McHugh, C.W., Ager, A.A., Bahro, B., Agee, J.K., 2007. Simulation of long-term landscape-level fuel treatment effects on large wildfires. *Int. J. Wildland Fire* 16, 712–727. <https://doi.org/10.1071/WF06064>

- Fischer, P.W., Cullen, A.C., Ettl, G.J., 2017. The Effect of Forest Management Strategy on Carbon Storage and Revenue in Western Washington: A Probabilistic Simulation of Tradeoffs. *Risk Anal.* 37, 173–192. <https://doi.org/10.1111/risa.12611>
- Foley, T.G., Richter, D. deB., Galik, C.S., 2009. Extending rotation age for carbon sequestration: A cross-protocol comparison of North American forest offsets. *For. Ecol. Manag.* 259, 201–209. <https://doi.org/10.1016/j.foreco.2009.10.014>
- Foody, G.M., Boyd, D.S., Cutler, M.E.J., 2003. Predictive relations of tropical forest biomass from Landsat TM data and their transferability between regions. *Remote Sens. Environ.* 85, 463–474. [https://doi.org/10.1016/S0034-4257\(03\)00039-7](https://doi.org/10.1016/S0034-4257(03)00039-7)
- Forest Climate Working Group, 2015. Forest Carbon Solutions for Mitigating Climate Change: A Toolkit for State Governments. American Forest Foundation, Washington, D.C.
- Franklin, J.F., Dyrness, C.T., 1988. Natural Vegetation of Oregon and Washington. Oregon State University Press.
- Fried, J.S., Jain, T.B., Loreno, S., Keefe, R.F., Bell, C.K., 2017. A framework for evaluating forest restoration alternatives and their outcomes, over time, to inform monitoring: Bioregional inventory originated simulation under management (Proceedings of the 2017 Forest Vegetation Simulator (FVS) Virtual e-Conference No. GTR-SRS-224). U.S. Department of Agriculture Forest Service, Southern Research Station., Asheville, NC.
- Froese, R.E., Robinson, A.P., 2007. A validation and evaluation of the Prognosis individual-tree basal area increment model. *Can. J. For. Res.* 37, 1438–1449. <https://doi.org/10.1139/X07-002>
- FSC International, 2015. FSC Principles and Criteria for Forest Stewardship (No. FSC-STD-01-001 V5-2 EN). Forest Stewardship Council, Bonn, Germany.
- FSC-US, 2012. FSC-US Forest Management Standard (v1.0). Forest Stewardship Council International.
- Fulé, P.Z., Cocke, A.E., Heinlein, T.A., Covington, W.W., 2004. Effects of an Intense Prescribed Forest Fire: Is It Ecological Restoration? *Restor. Ecol.* 12, 220–230. <https://doi.org/10.1111/j.1061-2971.2004.00283.x>
- FVS Staff, 2023a. Klamath Mountains (NC) Variant Overview of the Forest Vegetation Simulator (Internal Report). U. S. Department of Agriculture, Forest Service, Forest Management Service Center, Fort Collins, CO.
- FVS Staff, 2023b. Inland California and Southern Cascades (CA) Variant Overview of the Forest Vegetation Simulator (Internal Report). U. S. Department of Agriculture, Forest Service, Forest Management Service Center, Fort Collins, CO.
- FVS Staff, 2022a. Westside Cascades (WC) Variant Overview of the Forest Vegetation Simulator (Internal Report). U. S. Department of Agriculture, Forest Service, Forest Management Service Center, Fort Collins, CO.
- FVS Staff, 2022b. Pacific Northwest Coast (PN) Variant Overview of the Forest Vegetation Simulator (Internal Report). U. S. Department of Agriculture, Forest Service, Forest Management Service Center, Fort Collins, CO.
- Galik, C.S., Cooley, D.M., Baker, J.S., 2012. Analysis of the production and transaction costs of forest carbon offset projects in the USA. *J. Environ. Manage.* 112, 128–136. <https://doi.org/10.1016/j.jenvman.2012.06.045>
- Gálvez, F.B., Hudak, A.T., Byrne, J.C., Crookston, N.L., Keefe, R.F., 2014. Using climate-FVS to project landscape-level forest carbon stores for 100 years from field and LiDAR measures of initial conditions. *Carbon Balance Manag.* 9, 1.

- Gelman, A., Goodrich, B., Gabry, J., Vehtari, A., 2019. R-squared for Bayesian Regression Models. *Am. Stat.* 73, 307–309. <https://doi.org/10.1080/00031305.2018.1549100>
- Gelman, A., Vehtari, A., Simpson, D., Margossian, C.C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C., Modrák, M., 2020. Bayesian Workflow. <https://doi.org/10.48550/arXiv.2011.01808>
- Gerber, L.R., Barton, C.J., Cheng, S.H., Anderson, D., 2020. Producing actionable science in conservation: Best practices for organizations and individuals. *Conserv. Sci. Pract.* 2. <https://doi.org/10.1111/csp2.295>
- Gholz, H.L., 1982. Environmental Limits on Aboveground Net Primary Production, Leaf Area, and Biomass in Vegetation Zones of the Pacific Northwest. *Ecology* 63, 469–481. <https://doi.org/10.2307/1938964>
- Gilmore, D.W., 2003. To Thin or Not to Thin: Using the Forest Vegetation Simulator to Evaluate Thinning of Aspen. *North. J. Appl. For.* 20, 14–18. <https://doi.org/10.1093/njaf/20.1.14>
- Glasby, M.J., Russell, M.B., Domke, G.M., 2019. Analyzing the impacts of forest disturbance on individual tree diameter increment across the US Lake States. *Environ. Monit. Assess.* 191, 56. <https://doi.org/10.1007/s10661-019-7187-8>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ., Big Remotely Sensed Data: tools, applications and experiences* 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Gray, A.N., McIntosh, A.C.S., Garman, S.L., Shettles, M.A., 2021. Predicting canopy cover of diverse forest types from individual tree measurements. *For. Ecol. Manag.* 501, 119682. <https://doi.org/10.1016/j.foreco.2021.119682>
- Gunnoe, A., 2016. The Financialization of the US Forest Products Industry: Socio-Economic Relations, Shareholder Value, and the Restructuring of an Industry. *Soc. Forces* 94, 1075–1101. <https://doi.org/10.1093/sf/sov108>
- Haberland, C., Marston, J., 2017. Spatial Modeling of Timber Ecosystem Services: Linking the FVS Econ Extension and FSVeg Spatial Data Analyzer to Map Stumpage Value (General Technical Report No. SRS-GTR-224), Proceedings of the 2017 Forest Vegetation Simulator (FVS) e-Conference. U.S. Department of Agriculture Forest Service, Southern Research Station, Asheville, NC.
- Hamrick, K., Gallant, M., 2017. Fertile Ground: State of Forest Carbon Finance 2017. Forest Trends' Ecosystem Marketplace, Washington, D.C.
- Hanson, E.J., Azuma, D.L., Hiserote, B.A., 2002. Site index equations and mean annual increment equations for Pacific Northwest Research Station Forest Inventory and Analysis Inventories, 1985-2001 (Research Note No. PNW-RN-533). US Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR.
- Harmon, M.E., Marks, B., 2002. Effects of silvicultural practices on carbon stores in Douglas-fir – western hemlock forests in the Pacific Northwest, U.S.A.: results from a simulation model 32, 15.
- Haynes, R., 2005. Will Markets Provide Sufficient Incentive for Sustainable Forest Management?, in: Deal, R.L., White, S.M. (Eds.), *Understanding Key Issues of Sustainable Wood Production in the Pacific Northwest*. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR, pp. 13–19. <https://doi.org/10.2737/PNW-GTR-626>

- Haynes, R.W., 2003. An analysis of the timber situation in the United States: 1952 to 2050. (No. PNW-GTR-560). U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR. <https://doi.org/10.2737/PNW-GTR-560>
- Heilman, K.A., Dietze, M.C., Arizpe, A.A., Aragon, J., Gray, A., Shaw, J.D., Finley, A.O., Klesse, S., DeRose, R.J., Evans, M.E.K., 2022. Ecological forecasting of tree growth: Regional fusion of tree-ring and forest inventory data to quantify drivers and characterize uncertainty. *Glob. Change Biol.* 28, 2442–2460. <https://doi.org/10.1111/gcb.16038>
- Herbert, C., Fried, J.S., Butsic, V., 2023. Validation of Forest Vegetation Simulator Model Finds Overprediction of Carbon Growth in California. *Forests* 14, 604. <https://doi.org/10.3390/f14030604>
- Herbold, S., 2020. Autorank: A Python package for automated ranking of classifiers. *J. Open Source Softw.* 5, 2173. <https://doi.org/10.21105/joss.02173>
- Hooper, S., Kennedy, R.E., 2018. A spatial ensemble approach for broad-area mapping of land surface properties. *Remote Sens. Environ.* 210, 473–489. <https://doi.org/10.1016/j.rse.2018.03.032>
- Hoover, C.M., Bagdon, B., Gagnon, A., 2021. Standard estimates of forest ecosystem carbon for forest types of the United States (General Technical Report No. NRS-GTR-202). U.S. Department of Agriculture, Forest Service, Northern Research Station, Madison, WI. <https://doi.org/10.2737/NRS-GTR-202>
- Hudak, A.T., Fekety, P.A., Kane, V.R., Kennedy, R.E., Filippelli, S.K., Falkowski, M.J., Tinkham, W.T., Smith, A.M.S., Crookston, N.L., Domke, G.M., Corrao, M.V., Bright, B.C., Churchill, D.J., Gould, P.J., McGaughey, R.J., Kane, J.T., Dong, J., 2020. A carbon monitoring system for mapping regional, annual aboveground biomass across the northwestern USA. *Environ. Res. Lett.* 15, 095003. <https://doi.org/10.1088/1748-9326/ab93f9>
- Hudak, A.T., Haren, A.T., Crookston, N.L., Liebermann, R.J., Ohmann, J.L., 2014. Imputing Forest Structure Attributes from Stand Inventory and Remotely Sensed Data in Western Oregon, USA. *For. Sci.* 60, 253–269. <https://doi.org/10.5849/forsci.12-101>
- Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* 25, 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)
- Jin, S., Su, Y., Gao, S., Hu, T., Liu, J., Guo, Q., 2018. The Transferability of Random Forest in Canopy Height Estimation from Multi-Source Remote Sensing Data. *Remote Sens.* 10, 1183. <https://doi.org/10.3390/rs10081183>
- Kennedy, R., Yang, Z., Gorelick, N., Braaten, J., Cavalcante, L., Cohen, W., Healey, S., 2018. Implementation of the LandTrendr Algorithm on Google Earth Engine. *Remote Sens.* 10, 691. <https://doi.org/10.3390/rs10050691>
- Kennedy, R.E., Yang, Z., Cohen, W.B., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr — Temporal segmentation algorithms. *Remote Sens. Environ.* 114, 2897–2910. <https://doi.org/10.1016/j.rse.2010.07.008>
- Kerchner, C.D., Keeton, W.S., 2015. California’s regulatory forest carbon market: Viability for northeast landowners. *For. Policy Econ.* 50, 70–81. <https://doi.org/10.1016/j.forpol.2014.09.005>
- Keyser, C., 2017. Pacific Northwest Coast (PN) Variant Overview – Forest Vegetation Simulator (Internal Report). U. S. Department of Agriculture, Forest Service, Forest Management Service Center, Fort Collins, CO.

- Kirchhoff, C.J., Carmen Lemos, M., Dessai, S., 2013. Actionable Knowledge for Environmental Decision Making: Broadening the Usability of Climate Science. *Annu. Rev. Environ. Resour.* 38, 393–414. <https://doi.org/10.1146/annurev-environ-022112-112828>
- Klesse, S., DeRose, R.J., Guiterman, C.H., Lynch, A.M., O'Connor, C.D., Shaw, J.D., Evans, M.E.K., 2018. Sampling bias overestimates climate change impacts on forest growth in the southwestern United States. *Nat. Commun.* 9, 5336. <https://doi.org/10.1038/s41467-018-07800-y>
- Korzukhin, M.D., Ter-Mikaelian, M.T., Wagner, R.G., 1996. Process versus empirical models: which approach for forest ecosystem management? *Can. J. For. Res.* 26, 879–887. <https://doi.org/10.1139/x26-096>
- Krajicek, J.E., Brinkman, K.A., Gingrich, S.F., 1961. Crown Competition - A Measure of Density. *For. Sci.* 7, 35–42.
- Krstajic, D., Buturovic, L.J., Leahy, D.E., Thomas, S., 2014. Cross-validation pitfalls when selecting and assessing regression and classification models. *J. Cheminformatics* 6, 10. <https://doi.org/10.1186/1758-2946-6-10>
- Kruschke, J., 2014. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, 2nd edition. ed. Academic Press, Boston.
- Kruschke, J.K., 2018. Rejecting or Accepting Parameter Values in Bayesian Estimation. *Adv. Methods Pract. Psychol. Sci.* 1, 270–280. <https://doi.org/10.1177/2515245918771304>
- Kruschke, J.K., Liddell, T.M., 2018. *The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective*. *Psychon. Bull. Rev.* 25. <https://doi.org/10.3758/s13423-016-1221-4>
- Landsberg, J., 2003. Modelling forest ecosystems: state of the art, challenges, and future directions. *Can. J. For. Res.* 33, 385–397. <https://doi.org/10.1139/x02-129>
- Latta, G., Temesgen, H., Barrett, T.R., 2009. Mapping and imputing potential productivity of Pacific Northwest forests using climate variables. *Can. J. For. Res.* 39, 1197–1207. <https://doi.org/10.1139/X09-046>
- Law, B.E., Hudiburg, T.W., Berner, L.T., Kent, J.J., Buotte, P.C., Harmon, M.E., 2018. Land use strategies to mitigate climate change in carbon dense temperate forests. *Proc. Natl. Acad. Sci.* 115, 3663–3668. <https://doi.org/10.1073/pnas.1720064115>
- LEMMA, 2014. GNN Plot Database [WWW Document]. URL <https://lemma.forestry.oregonstate.edu/data/plot-database> (accessed 5.16.18).
- Loog, M., 2012. Nearest neighbor-based importance weighting, in: 2012 IEEE International Workshop on Machine Learning for Signal Processing. Presented at the 2012 IEEE International Workshop on Machine Learning for Signal Processing, pp. 1–6. <https://doi.org/10.1109/MLSP.2012.6349714>
- Luther, J.E., Fournier, R.A., van Lier, O.R., Bujold, M., 2019. Extending ALS-Based Mapping of Forest Attributes with Medium Resolution Satellite and Environmental Data. *Remote Sens.* 11, 1092. <https://doi.org/10.3390/rs11091092>
- Maffei, H., Tandy, B., 2002. Methodology for Modeling the Spatial and Temporal Effects of Vegetation Management Alternatives on Late Successional Habitat in the Pacific Northwest (Proceedings of the Second Forest Vegetation Simulator Conference, February 2002 No. RMRS-P-25). U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Ogden, UT.

- Marino, E., Tomé, J.L., Hernando, C., Guijarro, M., Madrigal, J., 2022. Transferability of Airborne LiDAR Data for Canopy Fuel Mapping: Effect of Pulse Density and Model Formulation. *Fire* 5, 126. <https://doi.org/10.3390/fire5050126>
- Martin, F.C., 2013. User Guide to the Economic Extension (ECON) of the Forest Vegetation Simulator (Internal Report). U.S. Department of Agriculture, Forest Service, Forest Management Service Center, Fort Collins, CO.
- Mauro, F., Hudak, A.T., Fekety, P.A., Frank, B., Temesgen, H., Bell, D.M., Gregory, M.J., McCarley, T.R., 2021. Regional Modeling of Forest Fuels and Structural Attributes Using Airborne Laser Scanning Data in Oregon. *Remote Sens.* 13, 261. <https://doi.org/10.3390/rs13020261>
- McArdle, R.E., Meyer, W.H., Bruce, D., 1961. The Yield of Douglas Fir in the Pacific Northwest (Technical Bulletin No. 201). USDA Forest Service, Pacific Northwest Forest and Range Experiment Station, Washington, D.C.
- McGaughey, R.J., 2022. FUSION/LDV: Software for LIDAR Data Analysis and Visualization.
- McGaughey, R.J., 2004. Stand Visualization System. USDA Forest Service, Pacific Northwest Research Station, Portland, OR.
- McKelvey, K.S., Block, W.M., Jain, T.B., Luce, C.H., Page-Dumroese, D.S., Richardson, B.A., Saab, V.A., Schoettle, A.W., Sieg, C.H., Williams, D.R., 2021. Adapting Research, Management, and Governance to Confront Socioecological Uncertainties in Novel Ecosystems. *Front. For. Glob. Change* 4, 14. <https://doi.org/10.3389/ffgc.2021.644696>
- Melson, S., Azuma, D., Fried, J.S., 2002. A first look at measurement error on FIA plots using blind plots in the Pacific Northwest, in: *Proceedings of the Third Annual Forest Inventory and Analysis Symposium*. U.S. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN, pp. 11–20.
- Mendell, B., Lang, A.H., 2013. Comparing Forest Certification Standards in the U.S.: Economic Analysis and Practical Considerations. *EconoSTATS*, Fairfax, VA.
- Mitchell, K.J., Cameron, I.R., 1985. Managed stand yield tables for coastal Douglas-fir: initial density and precommercial thinning (Land Management Report No. 31). British Columbia Ministry of Forests, Research Branch, Victoria, B.C.
- NASA, 2000. Measuring Vegetation (NDVI & EVI) [WWW Document]. NASA Earth Obs. URL https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetation_2.php (accessed 8.15.23).
- Nebel, G., Quevedo, L., Bredahl Jacobsen, J., Helles, F., 2005. Development and economic significance of forest certification: the case of FSC in Bolivia. *For. Policy Econ.* 7, 175–186. [https://doi.org/10.1016/S1389-9341\(03\)00030-3](https://doi.org/10.1016/S1389-9341(03)00030-3)
- Nemenyi, P., 1963. Distribution-free Multiple Comparisons (Ph.D. Dissertation). Princeton University, Princeton, N.J.
- New Forests, 2013. Timberland Investment Outlook 2013-2017. New Forests, Chatswood, NSW, Australia.
- Ogle, K., Barber, J.J., 2020. Ensuring identifiability in hierarchical mixed effects Bayesian models. *Ecol. Appl.* 30, e02159. <https://doi.org/10.1002/eap.2159>
- Ohmann, J.L., Gregory, M.J., 2002. Predictive mapping of forest composition and structure with direct gradient analysis and nearest- neighbor imputation in coastal Oregon, U.S.A. *Can. J. For. Res.* 32, 725–741. <https://doi.org/10.1139/x02-011>

- Ohmann, J.L., Gregory, M.J., Henderson, E.B., Roberts, H.M., 2011. Mapping gradients of community composition with nearest-neighbour imputation: extending plot data for landscape analysis: Extending plot data for landscape analysis. *J. Veg. Sci.* 22, 660–676. <https://doi.org/10.1111/j.1654-1103.2010.01244.x>
- Oregon Forest Resources Institute, 2017. Oregon Forest Facts: 2017-18 Edition. Oregon Forest Resources Institute, Portland, OR.
- Oregon Secretary of State, n.d. Oregon Forest Practice Rules, Oregon Administrative Rules.
- Papaspiliopoulos, O., Roberts, G.O., Sköld, M., 2003. Non-Centered Parameterisations for Hierarchical Models and Data Augmentation. *Bayesian Stat.* 7, 307–326.
- Pearman, O., Cravens, A.E., 2022. Institutional barriers to actionable science: Perspectives from decision support tool creators. *Environ. Sci. Policy* 128, 317–325. <https://doi.org/10.1016/j.envsci.2021.12.004>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Phan, D., Pradhan, N., Jankowiak, M., 2019. Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro. <https://doi.org/10.48550/arXiv.1912.11554>
- Pinchot Institute for Conservation, 2011. Forest Carbon Incentives: Options for Landowner Incentives to Increase Forest Carbon Sequestration. Pinchot Institute for Conservation, Washington, D.C.
- Pokharel, B., Froese, R.E., 2008. Evaluating alternative implementations of the Lake States FVS diameter increment model. *For. Ecol. Manag.* 255, 1759–1771. <https://doi.org/10.1016/j.foreco.2007.11.035>
- Rabotyagov, S.S., Lin, S., 2013. Small forest landowner preferences for working forest conservation contract attributes: A case of Washington State, USA. *J. For. Econ.* 19, 307–330.
- Rabotyagov, S.S., Tóth, S.F., Ettl, G.J., 2013. Testing the Design Variables of ECOSEL: A Market Mechanism for Forest Ecosystem Services. *For. Sci.* 59, 303–321. <https://doi.org/10.5849/forsci.11-110>
- Renner, D.L., Martin, F., 2002. Using the Fuels and Fire Effects (FFE) and Economic (ECON) Extensions to the Forest Vegetation Simulator (FVS) to Evaluate the Impacts of Silvicultural Regimes (Proceedings of the Second Forest Vegetation Simulator Conference, February 2002 No. RMRS-P-25). U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Ogden, UT.
- Riemann, R., Wilson, B.T., Lister, A., Parks, S., 2010. An effective assessment protocol for continuous geospatial datasets of forest characteristics using USFS Forest Inventory and Analysis (FIA) data. *Remote Sens. Environ.* 114, 2337–2352. <https://doi.org/10.1016/j.rse.2010.05.010>
- Roberts, J.C., 2002. Using a Multichange Agent Approach with the Forest Vegetation Simulator on the Boise National Forest, Idaho (Proceedings of the Second Forest Vegetation Simulator Conference, February 2002 No. RMRS-P-25). U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Ogden, UT.
- Russell, M.B., D’Amato, A.W., Albers, M.A., Woodall, C.W., Puettmann, K.J., Saunders, M.R., VanderSchaaf, C.L., 2015. Performance of the Forest Vegetation Simulator in Managed

- White Spruce Plantations Influenced by Eastern Spruce Budworm in Northern Minnesota. *For. Sci.* 61, 723–730. <https://doi.org/10.5849/forsci.14-150>
- Russell, M.B., Weiskittel, A.R., Kershaw, J.A., 2013. Benchmarking and Calibration of Forest Vegetation Simulator Individual Tree Attribute Predictions Across the Northeastern United States. *North. J. Appl. For.* 30, 75–84. <https://doi.org/10.5849/njaf.12-034>
- Sánchez Meador, A.J., Waring, K.M., Kalies, E.L., 2015. Implications of Diameter Caps on Multiple Forest Resource Responses in the Context of the Four Forests Restoration Initiative: Results from the Forest Vegetation Simulator. *J. For.* 113, 219–230. <https://doi.org/10.5849/jof.14-021>
- Sathre, R., O'Connor, J., 2010. Meta-analysis of greenhouse gas displacement factors of wood product substitution. *Environ. Sci. Policy* 13, 104–114. <https://doi.org/10.1016/j.envsci.2009.12.005>
- Schliep, E.M., Dong, T.Q., Gelfand, A.E., Li, F., 2014. Modeling individual tree growth by fusing diameter tape and increment core data. *Environmetrics* 25, 610–620. <https://doi.org/10.1002/env.2324>
- Schroder, S.A. (Kushch), Toth, S.F., Deal, R.L., Ettl, G.J., 2016. Multi-objective optimization to evaluate tradeoffs among forest ecosystem services following fire hazard reduction in the Deschutes National Forest, USA. *Ecosyst. Serv.* 22, 328–347. <https://doi.org/10.1016/j.ecoser.2016.08.006>
- Schumacher, F.X., 1930. Yield, stand and volume tables for Douglas fir in California (Bulletin No. 491). University of California, Berkeley, Agricultural Experiment Station, Berkeley, CA.
- Shaw, J.D., 2009. Using FIA Data in the Forest Vegetation Simulator, in: USDA Forest Service Proceedings - RMRS-P-56. Presented at the 2008 Forest Inventory and Analysis (FIA) Symposium; October 21-23, 2008: Park City, UT, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO, p. 16.
- Shaw, J.D., Gagnon, A., 2020a. Quick-Start Guide to Forest Inventory and Analysis Data in the Forest Vegetation Simulator. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO.
- Shaw, J.D., Gagnon, A., 2020b. Field Note: A New Conversion of Forest Inventory and Analysis Data for Use in the Forest Vegetation Simulator. *J. For.* 118, 307–312. <https://doi.org/10.1093/jofore/fvz050>
- Shaw, J.D., Vacchiano, G., Brough, A., Kusbach, A., Long, J.N., 2006. Local calibration of the Forest Vegetation simulator (FVS) using custom inventory data, in: Proceedings. Presented at the Society of American Foresters 2006 National Convention, Society of American Foresters, Pittsburgh, PA, p. 12.
- Shi, T., Xu, H., 2019. Derivation of Tasseled Cap Transformation Coefficients for Sentinel-2 MSI At-Sensor Reflectance Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 12, 4038–4048. <https://doi.org/10.1109/JSTARS.2019.2938388>
- Sohnngen, B., Brown, S., 2008. Extending timber rotations: carbon and cost implications. *Clim. Policy* 8, 435–451.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., van der Linde, A., 2014. The deviance information criterion: 12 years on. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 76, 485–493.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 64, 583–639. <https://doi.org/10.1111/1467-9868.00353>

- Stand Management Cooperative, 2016. SMC Plantation Yield Calculator [WWW Document]. Stand Manag. Coop. URL <http://www.sefs.washington.edu/research.smc/research/pyc/index.html> (accessed 5.3.18).
- Strunk, J., Packalen, P., Gould, P., Gatzolis, D., Maki, C., Andersen, H.-E., McGaughey, R.J., 2019. Large area forest yield estimation with pushbroom Digital Aerial Photogrammetry. *For.* 105 397– 10, 397. <https://doi.org/10.3390/f10050397>
- Talbert, C., Marshall, D., 2005. Plantation Productivity in the Douglas-Fir Region Under Intensive Silvicultural Practices: Results from Research and Operations. *J. For.* 103, 65–70. <https://doi.org/10.1093/jof/103.2.65>
- Tarancón, A.A., Fulé, P.Z., Shive, K.L., Sieg, C.H., Meador, A.S., Strom, B., 2014. Simulating post-wildfire forest trajectories under alternative climate and management scenarios. *Ecol. Appl.* 24, 1626–1637. <https://doi.org/10.1890/13-1787.1>
- Temesgen, H., Mauro, F., Hudak, A.T., Frank, B., Monleon, V., Fekety, P., Palmer, M., Bryant, T., 2021. Using Fay–Herriot Models and Variable Radius Plot Data to Develop a Stand-Level Inventory and Update a Prior Inventory in the Western Cascades, OR, United States. *Front. For. Glob. Change* 4.
- Thomas, K., Benjamin, R.-K., Fernando, P., Brian, G., Matthias, B., Jonathan, F., Kyle, K., Jessica, H., Jason, G., Sylvain, C., Paul, I., Damián, A., Safia, A., Carol, W., Team, J.D., 2016. Jupyter Notebooks - a publishing format for reproducible computational workflows, in: *Proceedings of the 20th International Conference on Electronic Publishing. Presented at the Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pp. 87–90. <https://doi.org/10.3233/978-1-61499-649-1-87>
- Tompalski, P., White, J.C., Coops, N.C., Wulder, M.A., 2019. Demonstrating the transferability of forest inventory attribute models derived using airborne laser scanning data. *Remote Sens. Environ.* 227, 110–124. <https://doi.org/10.1016/j.rse.2019.04.006>
- Tóth, S.F., Ettl, G.J., Könnnyű, N., Rabotyagov, S.S., Rogers, L.W., Cornick, J.M., 2013. ECOSEL: Multi-objective optimization to sell forest ecosystem services. *For. Policy Econ., New Frontiers of Forest Economics* 35, 73–82. <https://doi.org/10.1016/j.forpol.2013.06.011>
- Tóth, S.F., Ettl, G.J., Rabotyagov, S.S., 2010. ECOSEL: An Auction Mechanism for Forest Ecosystem Services. *Math. Comput. For. Nat.-Resour. Sci. MCFNS* 2, 99–116 (18).
- Tsamardinos, I., Rakhshani, A., Lagani, V., 2015. Performance-Estimation Properties of Cross-Validation-Based Protocols with Simultaneous Hyper-Parameter Optimization. *Int. J. Artif. Intell. Tools* 24, 1540023. <https://doi.org/10.1142/S0218213015400230>
- Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8, 127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)
- US Forest Carbon Modeling Group, in preparation. Natural Climate Solutions: Forest Modeling Research and Development Draft Priorities Report.
- US Geological Survey, 2022. Landsat Soil Adjusted Vegetation Index [WWW Document]. Landsat Missions. URL <https://www.usgs.gov/landsat-missions/landsat-soil-adjusted-vegetation-index> (accessed 8.15.23).
- USDA Forest Service, Forest Management Service Center, 2015. open-fvs: The Forest Vegetation Simulator (FVS) forest growth model [WWW Document]. Sourceforge. URL <https://sourceforge.net/projects/open-fvs/>

- USDA Forest Service, Natural Resource Manager, 2021. FS Veg/Common Stand Exam, FS Veg Spatial, and FS Veg Spatial Data Analyzer. National Forest Foundation, Missoula, MT.
- USDA Forest Service Pacific Northwest Research Station, 2017. PNW-FIADB: Pacific Northwest Annual Forest Inventory Database [WWW Document]. FIA Inventory Data. URL <https://www.fs.fed.us/pnw/rma/fia-topics/inventory-data/> (accessed 6.13.18).
- USFS FMSC, 2023. Forest Vegetation Simulator.
- USFS Forest Inventory and Analysis, 2023. Field Instructions for the Annual Inventory of California, Oregon, and Washington 2023. USDA Forest Service, Pacific Northwest Research Station, Resource Monitoring and Assessment Program, Portland, OR.
- Vacchiano, G., Shaw, J.D., DeRose, R.J., Long, J.N., 2008. Inventory-Based Sensitivity Analysis of the Large Tree Diameter Growth Submodel of the Southern Variant of the Forest Vegetation Simulator (Proceedings of the Third Forest Vegetation Simulator Conference, February 2007 No. RMRS-P-2008). U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO.
- van Kooten, G.C., Binkley, C.S., Delcourt, G., 1995. Effect of Carbon Taxes and Subsidies on Optimal Forest Rotation Age and Supply of Carbon Services. *Am. J. Agric. Econ.* 77, 365. <https://doi.org/10.2307/1243546>
- van Mantgem, P.J., Stephenson, N.L., Byrne, J.C., Daniels, L.D., Franklin, J.F., Fule, P.Z., Harmon, M.E., Larson, A.J., Smith, J.M., Taylor, A.H., Veblen, T.T., 2009. Widespread Increase of Tree Mortality Rates in the Western United States. *Science* 323, 521–524. <https://doi.org/10.1126/science.1165000>
- van Oijen, M., 2017. Bayesian Methods for Quantifying and Reducing Uncertainty and Error in Forest Models. *Curr. For. Rep.* 3, 269–280. <https://doi.org/10.1007/s40725-017-0069-9>
- Vandendriesche, D., 2010. FVS out of the box—assembly required, in: Proceedings RMRS-P-61. Presented at the 2009 National Silviculture Workshop (June 15-18, 2009), U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO, pp. 289–306.
- Varma, S., Simon, R., 2006. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 7, 91. <https://doi.org/10.1186/1471-2105-7-91>
- Vehtari, A., Gelman, A., Gabry, J., 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* 27, 1413–1432. <https://doi.org/10.1007/s11222-016-9696-4>
- Verra, 2012. Tool for the Demonstration and Assessment of Additionality in VCS Agriculture, Forestry and Other Land Use (AFOLU) Project Activities (version 3.0) (VCS Tool No. VT0001). Verified Carbon Standard, Washington, D.C.
- von Hagen, B., 2009. Unexplored Potential of Pacific Northwest Forests, in: *Old Growth in a New World: A Pacific Northwest Icon Reexamined*. Island Press, Washington, D.C., pp. 286–299.
- Waddell, K.L., Campbell, K., Kuegler, O., Christensen, G., 2014. FIA Volume Equation documentation updated on 9-19-2014.
- Wainer, J., Cawley, G., 2018. Nested cross-validation when selecting classifiers is overzealous for most practical applications. *ArXiv180909446 Cs Stat.*
- Wang, Y., 2017. Volume Estimator Library Equations. USDA Forest Service, Forest Management Service Center, Fort Collins, CO.
- Waring, R.H., Franklin, J.F., 1979. Evergreen Coniferous Forests of the Pacific Northwest. *Science* 204, 1380–1386.

- Washington Department of Natural Resources, 2018. Mill log prices - Domestically Processed (2/28/2018) [WWW Document]. Timber Sale QueryLog Prices. URL http://www.dnr.wa.gov/publications/psl_ts_feb18_logprices.pdf (accessed 6.13.18).
- Washington State Department of Natural Resources, 2009. Desired Future Condition Worksheet (Version 3.0) [WWW Document]. For. Pract. Appl. Rev. Syst. URL <https://fortress.wa.gov/dnr/protection/dfc/DfcRun.aspx> (accessed 6.27.18).
- Washington State Legislature, n.d. Washington State Forest Practices Rules, Washington Administrative Code.
- Weiskittel, A., Kershaw, J., Crookston, N., Hennigar, C., 2017. The Acadian Variant of the Forest Vegetation Simulator: Continued Development and Evaluation, in: Proceedings of the 2017 Forest Vegetation Simulator (FVS) e-Conference. U.S. Department of Agriculture Forest Service, Southern Research Station, Asheville, NC, pp. 10–13.
- Weiskittel, A.R., Garber, S.M., Johnson, G.P., Maguire, D.A., Monserud, R.A., 2007. Annualized diameter and height growth equations for Pacific Northwest plantation-grown Douglas-fir, western hemlock, and red alder. *For. Ecol. Manag.* 250, 266–278. <https://doi.org/10.1016/j.foreco.2007.05.026>
- Weiskittel, A.R., Maguire, D.A., Monserud, R.A., 2009. Development of a Hybrid Model for Intensively Managed Douglas-fir in the Pacific Northwest, in: Forest Growth and Timber Quality. Presented at the : Crown models and simulation methods for sustainable forest management (Portland, OR August 7-10, 2007), U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station., Portland, OR, pp. 49–67.
- Williamson, R.L., 1963. Growth and yield records from well-stocked stands of Douglas-fir (Research Paper No. PNW-RP-4). U.S. Department of Agriculture, Forest Service, Pacific Northwest Forest and Range Experiment Station, Portland, OR.
- Wilson, D., Monleon, V., Weiskittel, A., 2019. Quantification and Incorporation of Uncertainty in Forest Growth and Yield Projections Using A Bayesian Probabilistic Framework: A Demonstration for Plantation Coastal Douglas-fir in the Pacific Northwest, USA. *Math. Comput. For. Nat.-Resour. Sci. MCFNS* 11, 264-285(22).
- Woods, M.E., Robinson, D.C.E., 2008. Development of FVSOntario: A Forest Vegetation Simulator Variant and application software for Ontario, in: Proceedings of the Third Forest Vegetation Simulator Conference (February 13-15, 2007). U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO, pp. 212–224.
- Wykoff, W.R., 1990. A Basal Area Increment Model for Individual Conifers in the Northern Rocky Mountains. *For. Sci.* 36, 1077–1104.
- Zielinski, W.J., Gray, A.N., Dunk, J.R., Sherlock, J.W., Dixon, G.E., 2010. Using forest inventory and analysis data and the forest vegetation simulator to predict and monitor fisher (Martes pennanti) resting habitat suitability. Gen Tech Rep PSW-GTR-232 Albany CA US Dep. Agric. For. Serv. Pac. Southwest Res. Stn. 31 P. <https://doi.org/10.2737/PSW-GTR-232>
- Zou, H., Hastie, T., 2005. Regularization and Variable Selection Via the Elastic Net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 67, 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>