Forest Ecology and Management 409 (2018) 667–682

Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

Development and assessment of regeneration imputation models for National Forests of Oregon and Washington

Karin Kralicek[a], Andrew J. Sánchez Meador[b,*], Leah C. Rathbun[c]

[a] Oregon State University, Department of Forest Engineering, Resources, and Management, 210 Snell Hall, Corvallis, OR 97331, United States
[b] Northern Arizona University, School of Forestry, PO Box: 15018, Flagstaff, AZ 86011, United States
[c] USDA Forest Service, Pacific Northwest Regional Office, 1220 SW 3rd Ave, Portland, OR 97204, United States

ARTICLE INFO

Keywords:
Most similar neighbor
Forest inventory and analysis
Plant associations
Pacific Northwest
Multi-species

ABSTRACT

Imputation models were developed to predict seedling regeneration density and composition on National Forest System (NFS) lands in Oregon and Washington. The models were based on Forest Inventory and Analysis and Pacific Northwest Regional NFS Monitoring data. Individual models were developed based on broad forest plant association groups (FPAGs) with all model development and analysis conducted in R using a most similar neighbor-like imputation approach. Model performance was evaluated based on bias, mean absolute deviation, root mean-squared error (RMSE), and error rate in correctly predicting the total presence or absence of any regenerating species regardless of species (Total ER). Low to moderate RMSE ($\leq$7400 regeneration stems ha^{-1}) and low to moderate Total ER ($\leq$50%) were observed for 25 out of 58 FPAG-specific models. The regeneration imputation models produced in this study represent a large first step towards developing flexible, expandable, and adaptable regeneration models that can be easily incorporated into existing growth models like the Forest Vegetation Simulator.

1. Introduction

Seedling regeneration processes are integral components of forest stand dynamics and recovery following disturbance (Oliver and Larson, 1996, Crotteau et al., 2014) and can have significant influence on the outcome of growth model simulations, especially in forests dominated by natural regeneration (Weiskittel et al., 2011). Managers use these growth models to evaluate the potential effects of alternative treatments and these models are especially useful when site-specific information concerning treatment effects is lacking or unknown. However, not all growth models include regeneration components and many often leave this component to the user to provide as an input over time (Weiskittel et al., 2011). One reason for the lack of a regeneration component is that seedling regeneration (hereafter “regeneration”) is stochastic in nature and influenced by a wide variety of factors, many of which are not easily measured or observed (Weiskittel et al., 2011). It has also been noted that accurate prediction of regeneration can be complicated in areas composed of multiple tree species, numerous age cohorts, or with high structural complexity (Oliver and Larson, 1996, Ek et al., 1997). This is particularly true for the Forest Vegetation Simulator (FVS; Dixon, 2013), which is the standard growth model used by the US Forest Service and many other agencies throughout the United States and Canada.

The US Forest Service has developed regional variants of FVS across the United States and other organizations have developed regional variants for Canada (e.g., Lacerte et al., 2006). Yet, while FVS has become the growth model of choice for many public land managers, particularly federal land managers in the western Unites States, tools for automating natural regeneration remain limited or absent within the modeling process. At present, introducing natural regeneration automatically into an FVS simulation is only available for two geographic variants. For regional variants which cover the Northern Rocky Mountains, a model developed by Ferguson et al. (1986) and later extended by Ferguson and Carlson (1993) is available. This regeneration model predicts the expected density, size, species, and survival rate of seedlings following a disturbance, and makes use of stand characteristics that include site preparation methods and the residual species composition of the surrounding area. A similar model is available for coastal Alaska (Ferguson and Johnson, 1988). For all other geographic regions, the user must specify the density, size, species, and survival rate of expected seedlings.

At present, the user-specified approach is the only option available to managers using FVS in the Pacific Northwest where forests can range from single-species dominated stands to complex multi-cohort, multi-

* Corresponding author.
E-mail addresses: karin.kralicek@oregonstate.edu (K. Kralicek), andrew.sanchezmeador@nau.edu (A.J. Sánchez Meador), lrathbun@fs.fed.us (L.C. Rathbun).

https://doi.org/10.1016/j.foreco.2017.12.004
Received 3 July 2017; Received in revised form 29 November 2017; Accepted 2 December 2017
Available online 22 December 2017
0378-1127/

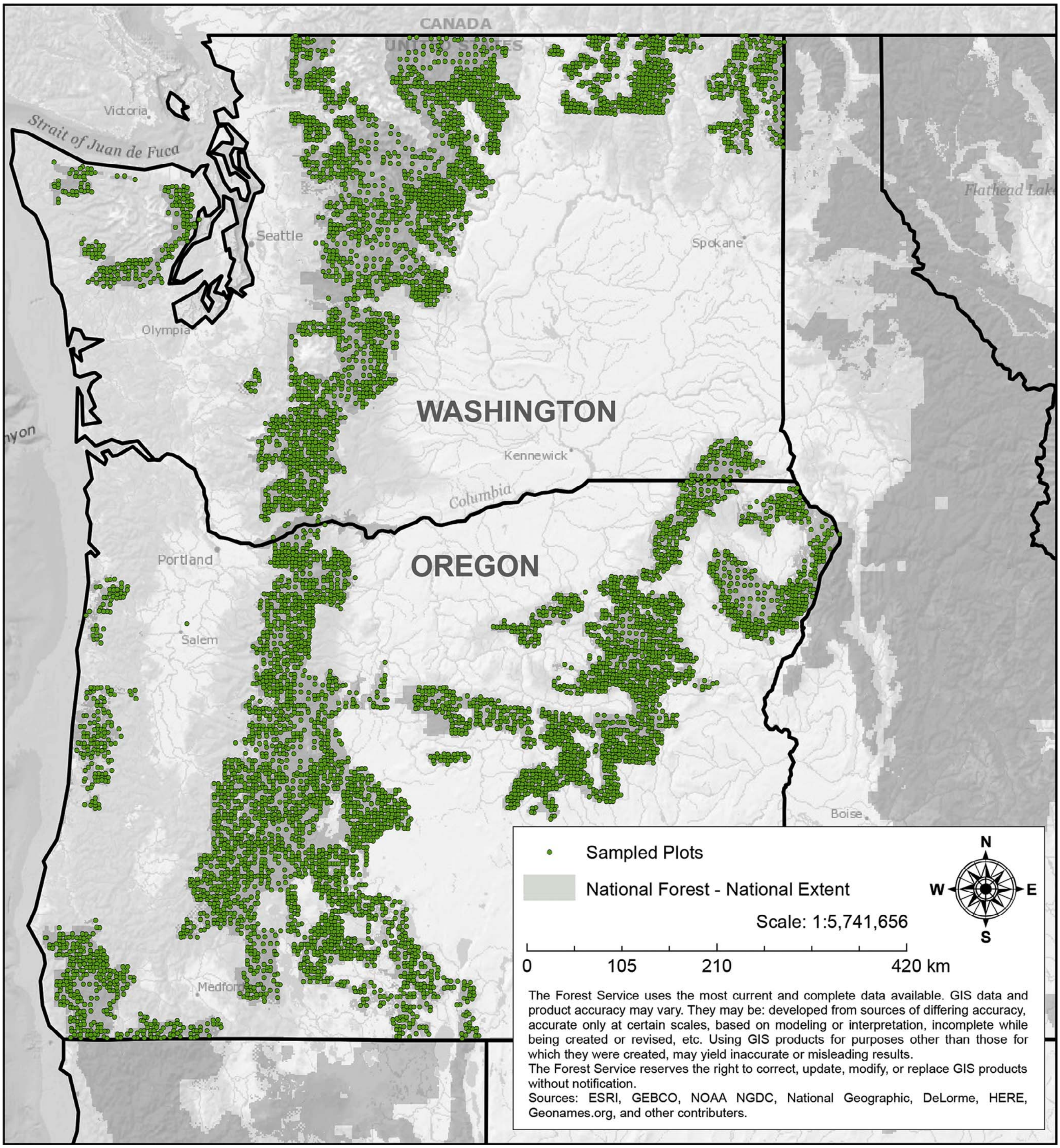

Fig. 1. Study area map illustrating the approximate location of all FIA and PNW subplots included in this study (green circles) from Oregon and Washington states in the Pacific Northwest region of the United States. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

species stands. The need exists for natural regeneration models that can be easily incorporated and automated into FVS, and that also cover the broad range of forested ecosystems found throughout the Pacific Northwest. Ideally, any models developed need to be implementable when data is restricted to the information most commonly collected by land managers; that is data captured in standard inventories (e.g. common stand exams) or permanent plots (e.g. Forest Inventory and Analysis, FIA).

Regeneration, and thus recruitment, can be modeled in many ways (see chapter 9 in Weiskittel et al., 2011 for a review), including regression based approaches, such as zero-inflated Poisson or negative binomial models. Recent examples where zero-inflated Poisson or negative binomial models have been successfully applied to model recruitment and regeneration include Fortin and DeBlois (2007) for forests of southern Québec, Li et al. (2011) for Acadian forests of southern Québec and Maine and the nearby study of Crotteau et al. (2014) for forests of northern California. Imputation methods are an alternative to classic regression techniques for modeling regeneration (Ek et al. 1997, Froese et al. (2003), Hassani et al. 2004). Imputation in this context is an estimation technique where missing or non-sampled measurements (such as regeneration) from one target stand are replaced with measurements or observations from another reference stand belonging to

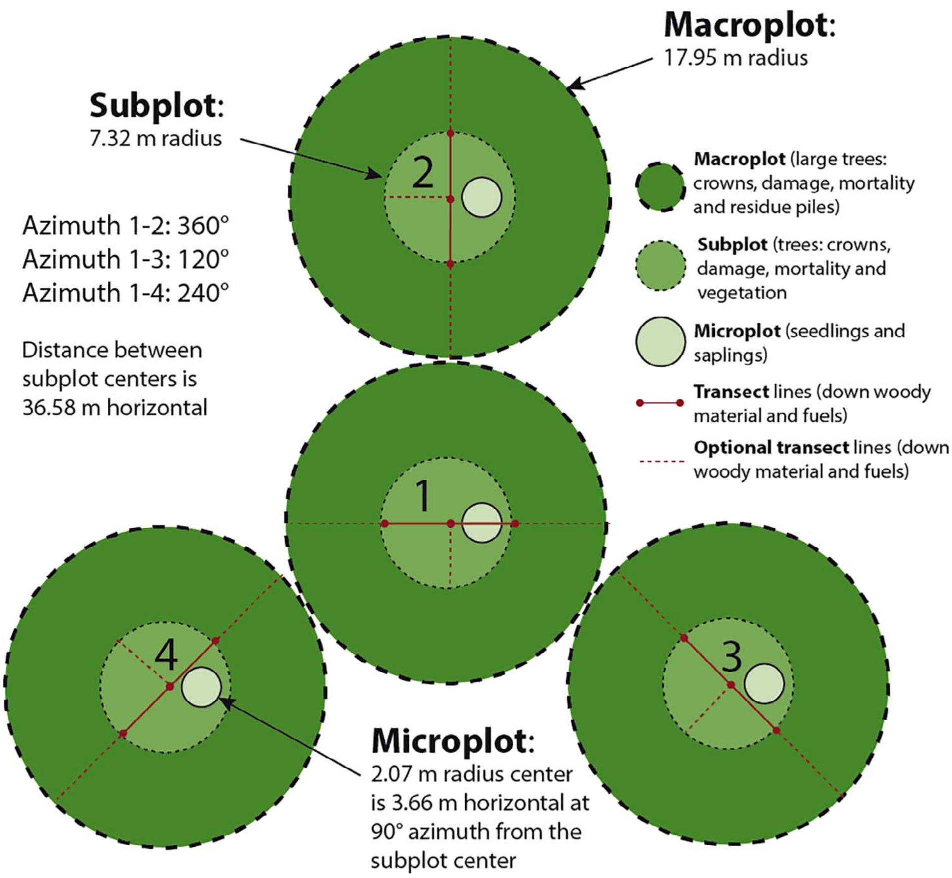

Fig. 2. The FIA mapped plot design. Subplot 1 is the center of the cluster with subplots 2, 3, and 4 located 36.58 m away at azimuths of 360°, 120°, and 240°, respectively. Regeneration is recorded on each microplot and overstory conditions are recorded on the subplot.

the same population and sharing similar characteristics with the target stand (Rubin, 1987, Ek et al., 1997, McRoberts, 2001, Eskelson et al., 2009b). Unlike regression alternatives, imputation methods do not require distributional assumptions about the data being modeled and can preserve associations between multiple stand attributes of interest (Moeur and Stage, 1995; McRoberts, 2001). In complex multi-species stands, imputation offers a venue for preserving naturally occurring species composition ratios. Furthermore, many imputation methods have the additional flexibility of improving predictions over time through the ease of incorporating new observations.

Two common methods used in the imputation of ecological data are most similar neighbor (MSN; e.g. Moeur and Stage, 1995) and gradient nearest neighbor (GNN; e.g. Ohmann and Gregory, 2002, and Ohmann et al., 2011). These methods differ in how they determine the *distance* to reference stands or *nearest neighbors*. MSN calculates distance metrics based on canonical correlation analysis, whereas GNN distances are based on canonical correspondence analysis (Eskelson et al., 2009a). With GNN and MSN, multiple similar nearest neighbors can be found for a single target stand. Generally if $k > 1$ nearest neighbors are selected, then an average of the k nearest neighbors are imputed to the original target stand. While imputing averages from $k > 1$ nearest neighbors can be attractive in that the likelihood of extreme values being imputed is down-weighted (the variability of the original dataset is preserved if $k = 1$), averages found from $k > 1$ nearest neighbors can produce predictions or combinations of imputed variables outside the bounds of reality (LeMay and Temesgen, 2005). Another approach to nearest neighbor selection is to first identify the $k > 1$ nearest neighbors and then, at random, select one of the k nearest neighbors for imputation. While this approach may not select the *closest* neighbor based on the distance method used, it has the advantage of producing results within the bounds of reality (by sampling from observed outcomes; no averaging) while also allowing for variability between imputed values for similar target stands (random selection).

The objectives of this study were to (1) explore the use of the imputation methods in developing natural regeneration models for the complex stands prevalent on lands of the Pacific Northwest Region of the U.S. Forest Service; and (2) develop and validate regeneration models that could be easily incorporated into FVS using commonly collected field data. In parallel with these overall objectives we wanted to develop flexible regeneration models that allow for multiple alternatives based upon empirical knowledge of a site. We also wanted to return model predictions that would include a random component to allow for differences in regeneration for similar sites.

2. Methods

2.1. Study area

Study area consisted of naturally regenerated National Forest System (NFS) lands in Oregon and Washington sampled between 2004 and 2013 (Fig. 1). NFS lands between these states encompass a broad range of vegetation types and climatic zones. Costal and dry-inland effects combine with the topographic effects of numerous mountain ranges and valleys to produce dramatic ranges in temperature and precipitation. This highly diverse region spans just over 10 million hectares in area (USFS, 2014) and contains a total of 846 plant associations (Hall, 1998). Coniferous forests dominate the study area with the most abundant tree species being Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco), western hemlock (*Tsuga heterophylla* (Raf.) Sarg.), ponderosa pine (*Pinus ponderosa* Lawson & C. Lawson), and lodgepole pine (*Pinus contorta* Douglas ex Loudon). Although areas dominated by a single species do exist on this landscape, many stands are characterized by complex assemblages of multiple species in the overstory with a potentially different composition of species regenerating underneath.

Table 1
Forest Inventory and Analysis intensive plot summary information for Forest Plant Association Groups with minimal adequate sample size (n ≥ 120) in Oregon and Washington state in the Pacific Northwest region of the United States. FPAGs are arranged in decreasing order by number of subplots available.

Forest Plant Association Group[A]	Code	Frequency of occurrence				Subplots				
		State	Nat. Forest	FVS variant[B]	Distinct Plant Assoc.[C]	Total	w/any Regen	w/mixed Regen	Mean[D] (Max.) Sp. Regen	Mean (Max.) Over. Sp
Ponderosa pine-Jeffrey pine/antelope bitterbrush-mountain mahogany	CPS2	2	10	2	16	1935	694	145	1.2 (4)	1.4 (4)
Pacific silver fir-noble fir/thinleaf huckleberry-rusty menziesia-Oregon boxleaf	CFS2	2	9	4	48	1900	1340	618	1.6 (7)	2.6 (7)
Douglas-fir/spiraea-common snowberry-kinnikinnick	CDS6	2	16	4	29	1585	413	94	1.3 (3)	1.8 (5)
Western hemlock/salal-Cascade barberry	CHS1	2	12	6	45	1565	596	178	1.4 (6)	2.4 (7)
Grand fir/pinegrass-Geyer's sedge	CWG1	2	10	3	12	1515	677	272	1.6 (6)	2.2 (6)
Douglas-fir/pinegrass-Geyer's sedge	CDG1	2	8	2	9	1254	434	131	1.4 (4)	1.8 (5)
Mountain hemlock/thinleaf huckleberry-rusty menziesia	CMS2	2	11	3	37	1197	955	494	1.8 (6)	2.6 (7)
Subalpine fir-Engelmann spruce (closed forest)	CE	2	11	2	21	1154	768	367	1.8 (6)	2.4 (6)
Douglas-fir/mallow ninebark	CDS7	2	5	1	4	858	178	23	1.2 (4)	1.7 (5)
Western hemlock/western swordfern/redwood-sorrel	CHF1	2	10	5	20	774	256	65	1.3 (4)	2.5 (7)
Ponderosa pine/grass-sedge	CPG2	2	8	2	5	748	254	53	1.2 (3)	1.4 (4)
Western hemlock/Pacific rhododendron	CHS3	2	7	5	26	704	286	94	1.5 (6)	2.5 (6)
White fir-grand fir/thinleaf huckleberry-creeping barberry	CWS2	2	10	3	11	687	362	165	1.7 (6)	2.4 (6)
White fir-grand fir/white spiraea-common snowberry (mesic)	CWS3	2	13	5	16	673	284	101	1.5 (5)	2.1 (6)
Ponderosa pine/bunchgrass (non pumice)	CPG1	2	9	2	6	566	145	32	1.3 (4)	1.3 (4)
White fir-grand fir/oceanspray-Cascade barberry-vine maple-salal	CWS5	2	13	6	17	554	215	91	1.6 (6)	2.4 (6)
Subalpine fir-spruce/twinflower	CEF2	2	7	1	14	522	314	129	1.6 (6)	2.4 (6)
Lodgepole pine/antelope bitterbrush (xeric)	CLS2	2	4	2	6	495	368	70	1.2 (4)	1.2 (3)
Mountain hemlock/grouse whortleberry/pinemat manzanita	CMS1	2	11	5	9	478	351	185	1.8 (6)	2.2 (7)
White fir-grand fir/twinflower	CWF3	2	12	4	7	460	293	159	2.0 (6)	2.4 (6)
White fir-grand fir/snowbrush ceanothus-greenleaf manzanita	CWS1	1	3	1	8	415	240	94	1.5 (4)	2.1 (6)
Western hemlockthinleaf huckleberry	CHS6	2	9	5	14	396	208	90	1.6 (5)	2.5 (5)
Tan oak-Canyon live oak	HTC6	1	1	2	7	395	89	19	1.3 (4)	2.6 (6)
Western hemlock/devilsclub	CHS5	2	8	4	12	380	65	28	1.5 (4)	2.1 (5)
Western hemlock/sweet after death-threeleaf foamflower (dry)	CHF2	2	9	4	7	362	146	45	1.4 (8)	2.4 (6)
White fir-grand fir/grouse huckleberry	CWS8	2	5	1	3	347	287	203	2.3 (6)	2.6 (6)
Lodgepole pine/grouse whortleberry	CLS4	2	7	2	9	339	260	129	1.8 (5)	1.9 (6)
Tanoak/California huckleberry	HTS1	1	1	2	7	338	43	8	1.2 (3)	2.5 (5)
White fir-grand fir/bride's bonnet/low forb	CWF4	2	9	4	11	336	146	61	1.7 (4)	2.4 (6)
Lodgepole pine/Idaho fescue (pumice)	CLG3	1	5	2	13	331	234	43	1.2 (4)	1.4 (4)
Douglas-fir/oceanspray-vine maple	CDS2	2	9	5	12	311	73	17	1.3 (3)	2 (6)
Ponderosa pine/common snowberry-rose spiraea	CPS5	2	7	2	5	287	74	11	1.2 (4)	1.4 (4)
Western hemlock/vine maple	CHS2	2	7	4	14	286	75	21	1.4 (4)	2.1 (6)
Pacific silver fir-noble fir/woodsorrel-twistedstalk-foamflower-bluebead	CFF1	2	9	4	8	280	150	57	1.5 (5)	2.8 (6)
Subalpine fir-spruce/rusty menziesia-rhododendron	CES2	2	7	1	5	272	212	97	1.7 (5)	2.3 (6)
Douglas-fir/chinquapin-canyon live oak	CDH5	2	7	5	10	268	74	22	1.5 (6)	2.2 (4)
Western red cedar/bride's bonnet	CCF221	2	2	2	1	266	164	59	1.6 (7)	3.5 (8)
Douglas-fir/blueberry	CDS8	2	6	1	7	246	141	51	1.5 (5)	2.1 (5)
Western hemlock/bride's bonnet-twinflower (mesic)	CHF3	2	3	2	3	235	164	82	1.9 (8)	3.8 (7)
Pacific silver fir-noble fir/Cascade barberry-salal	CFS1	2	5	3	6	223	136	58	1.5 (4)	3 (6)
Ponderosa pine/snowbrush ceonothus/Idaho fescue	CPS3	1	3	1	5	214	68	19	1.3 (3)	1.6 (4)
Douglas-fir/salal-creeping barberry	CDS5	1	5	4	5	207	62	22	1.5 (4)	2.2 (6)
Ponderosa pine/mountain big sagebrush/Geyer's sedge	CPS1	1	6	2	7	196	40	6	1.2 (2)	1.3 (3)
Pacific silver fir-noble fir/devilsclub	CFS3	2	6	3	3	191	115	49	1.5 (3)	2.3 (5)
Ponderosa pine/Wyethia	CPF1	1	3	1	3	173	62	11	1.2 (4)	1.5 (4)
Subalpine fir-spruce/Oregon boxleaf	CES1	2	6	1	4	173	86	27	1.4 (3)	2.2 (5)
White fir-grand fir-Shasta red fir-Pacific silver fir	CWC7	1	4	2	4	166	67	23	1.5 (4)	2.4 (5)
White fir-grand fir/mallow ninebark	CWS4	2	5	2	4	162	48	18	1.4 (3)	2.1 (6)

(continued on next page)

Table 1 (*continued*)

Forest Plant Association Group[A]	Code	Frequency of occurrence: State	Nat. Forest	FVS variant[B]	Distinct Plant Assoc.[C]	Subplots: Total	w/any Regen	w/mixed Regen	Mean[D] (Max.) Sp. Regen	Mean (Max.) Over. Sp
Douglas-fir/ceanothus-manzanita-Oregon boxleaf	CDS4	1	2	1	2	157	51	13	1.3 (3)	1.6 (5)
Douglas-fir-white fir	CDC421	1	4	2	1	152	60	37	2 (4)	2.6 (6)
Western juniper/mountain big sagebrush	CJS2	1	6	2	9	149	34	1	1 (2)	1.4 (4)
Douglas-fir	CD	2	15	7	1	144	57	23	1.5 (4)	2.1 (4)
Lodgepole pine/thinleaf huckleberry/pinegrass	CLS5	2	6	2	8	144	123	65	1.9 (7)	2.1 (5)
Shasta red fir/grouse whortleberry/pinemat manzanita	CRS1	1	4	2	2	142	103	53	1.7 (4)	2.2 (6)
Park	CAG1	2	9	3	12	141	79	34	1.5 (4)	1.9 (4)
Western juniper/bluebunch wheatgrass	CJG1	1	5	2	2	127	12	4	1.3 (2)	1.4 (4)
Subalpine fir-Engelmann spruce (wetlands)	CEM0	2	7	2	5	127	57	16	1.5 (6)	2.2 (6)
Subalpine fir-spruce/grass	CEG3	2	4	1	2	123	56	18	1.7 (4)	1.8 (5)

[A] As designated in Hall (1998) – available http://www.treesearch.fs.fed.us/pubs/5567.
[B] List of available FVS variants: Westside Cascades, Southeast Alaska and Coastal British Columbia, East Cascades, Inland Empire, Blue Mountains, South Central Oregon and Northeast California, and Pacific Northwest Coast.
[C] Number of distinct plant associations as designated in Hall (1998).
[D] Mean based on only subplots where regeneration was observed.

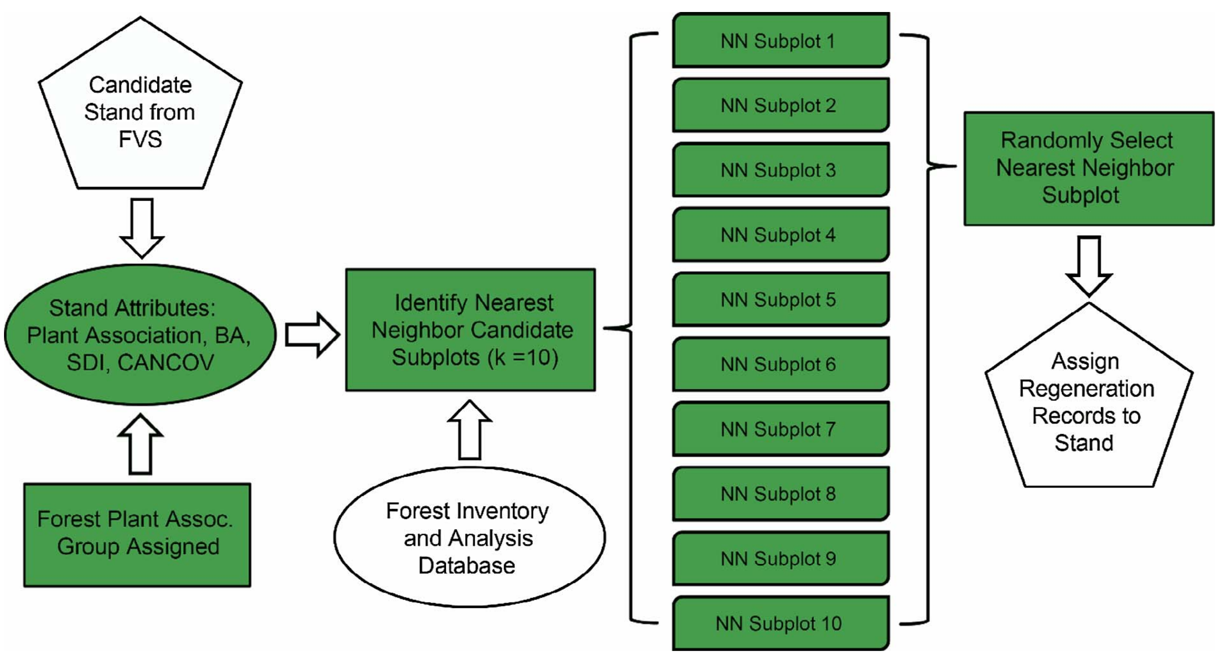

Fig. 3. Flow diagram of key model processes used in estimating regeneration.

2.2. Sample data and preparation

Forest Inventory and Analysis (FIA) data and Pacific Northwest (PNW) Regional NFS monitoring data from Oregon and Washington were accessed via the publically available FIA DataMart (http://fia-tools.fs.fed.us; USFS, 2013) and used as sample data. The FIA program is the national vegetation monitoring program in the United States and the PNW program was established by Region 6 of the USDA Forest Service as an intensification of FIA coverage on NFS land. FIA plots are randomly located across Oregon and Washington at a density of roughly one plot every 2428 hectares and, with the exception of Wilderness and Experimental Forests, there are three PNW plots for every one FIA plot.

The sampling protocol for both programs was similarly standardized in 2000, and the resulting plot design (Bechtold and Scott, 2005) is composed of one field plot that is a cluster of 4 points, each with an associated subplot, microplot, and macroplot (Fig. 2). Live trees were categorized into seedlings, saplings, and trees based on height and diameter. Seedlings were stems less than 2.54 cm in diameter and were measured on the microplot; to exclude first-year seedlings, a minimum length requirement of 0.15 m for conifers and 0.3 m for hardwood species was implemented. Saplings were stems between 2.54 and 12.5 cm in diameter at breast height (DBH; 1.37 m) and were also measured on the microplot; diameter for woodland species was taken at root collar, but for simplicity we will refer to all diameters as DBH. Trees were at least 12.5 cm in DBH and were measured on the subplot. The macroplot was used to capture infrequent observations of large trees, defined as having a minimum diameter of 76.2 or 60.96 cm for plots on the west or east side of the Cascade Range, respectively. Seedling counts were recorded by species on the microplot with diameter and height measured for all saplings and trees. Site information including plant association, fuels, slope, aspect, and condition with respect to structure, forest type, and management history were also collected at the subplot level.

Summary metrics and plant associations were aggregated at the subplot-level, instead of the plot-level, to retain specific combinations of species and count data and to preserve the scale at which plant associations were identified in the field. Aggregating at the subplot-level

Table 2
Number of FPAG-specific models by category of average RMSE and Total ER across 1000 simulations.

Average total RMSE (TPH) category[a]	Average total ER (%) category[b]			
	Low	Moderate	High	Total
Low	1	9	0	10
Moderate	1	14	0	15
High	1	31	1	33
Total	3	54	1	58

[A] RMSE category: low (≤3700 regeneration stems ha^{-1} (TPH)), moderate (3700–7400), high (≥7400).
[B] Total ER category: low (≤20%), moderate (20–40%), high (≥50%).

had the added benefit of avoiding loss of data that might occur by averaging across multiple subplots with drastically different structural characteristics or field-assigned plant associations (e.g., methods outlined for the Wenachee N.F. provided in Lillybridge et al. (1995)). To further ensure homogenous conditions with respect to regeneration, only subplots with their associated microplot dominated by a single site condition were included in the sample data. While sprouting species make up an important component of certain stands in the study area, modeling regeneration for such species fell outside the scope of this project due to insufficient data and time restrictions.

The Fia2Fvs utility (available: http://www.fs.fed.us/fmsc/fvs/software/data.shtml) was used to translate FIA data to an FVS readable Access database, after which FVS was used to generate potential prediction variables. To maximize accessibility to land managers with limited resources and data, only variables typical of a common stand exam and associated with plant association and summary metrics as calculated by FVS were considered as potential variables for the imputation models.

2.3. Imputation and model assessment

Due to the high number of plant associations in the study area (Fig. 1), non-distinct plant associations with respect to tree regeneration were generalized into broad forest plant associations groups (FPAG) for modeling purposes. Separate models were developed for each of the resulting 211 distinct FPAGs. While the majority of FPAGs (106) were composed of a single plant association, two FPAGs were composed of 40+ plant associations. Ten FPAGs had sample sizes over 1000 subplots, with the most common FPAG based on total number of subplots was the ponderosa – Jeffery pine/antelope bitterbrush – mountain mahogany (n = 1935). All FPAGs are represented in at least one variant of FVS, with three FPAGs represented in all seven variants currently developed for the Pacific Northwest Region.

However model evaluation was only preformed for FPAG-specific models that contained adequate sample observations, which was

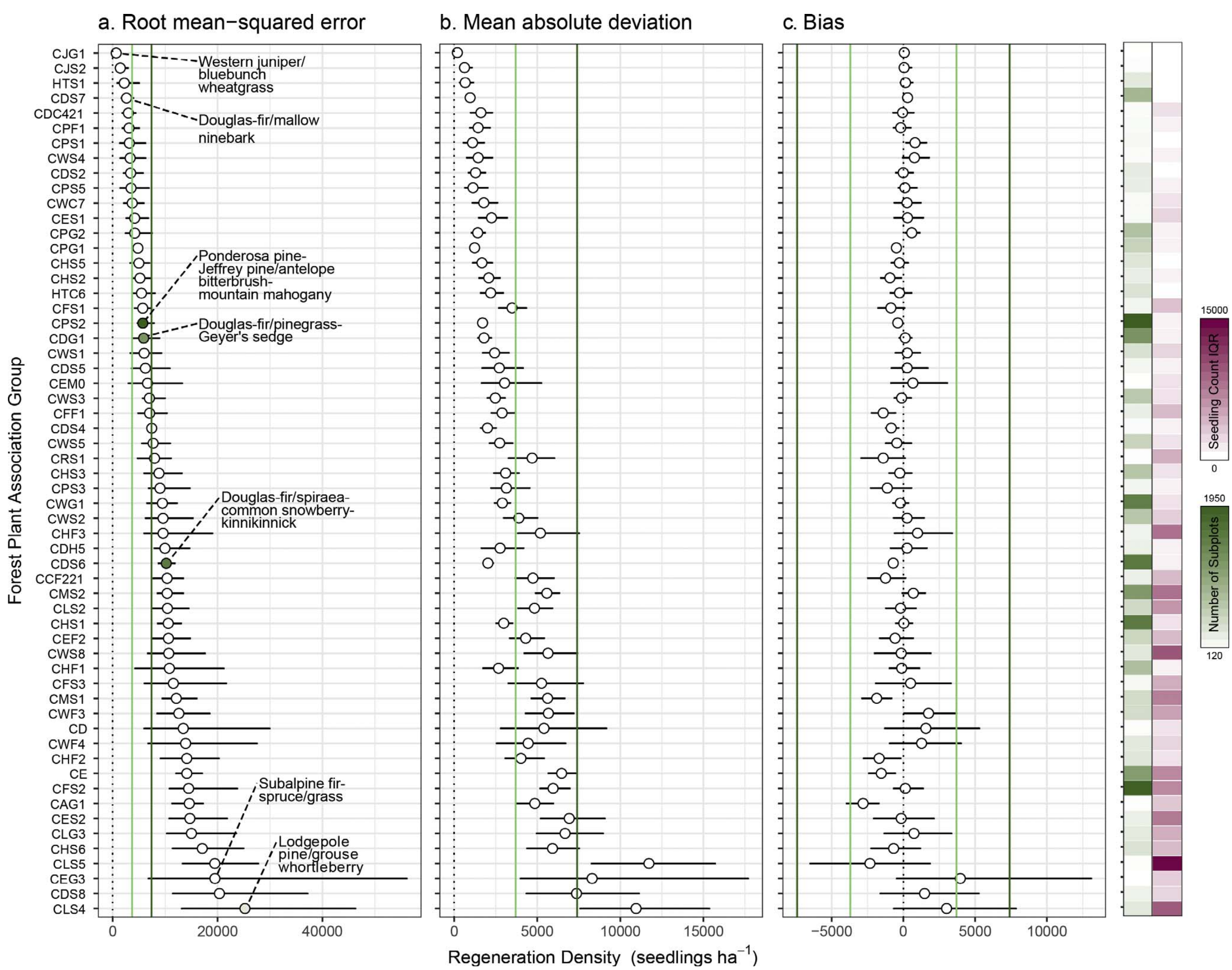

Fig. 4. Ninety-five percent confidence intervals based on 1000 simulations for average total (all species) (a) root mean-squared error (RMSE), (b) mean absolute deviation (MAD), and (c) bias. Summaries presented for models of Forest Plant Association Group with minimal adequate sample size (n ≥ 120) in Oregon and Washington state in the Pacific Northwest region of the United States. Number of subplots and observed interquartile range (IQR) for seedling densities are provided for reference.

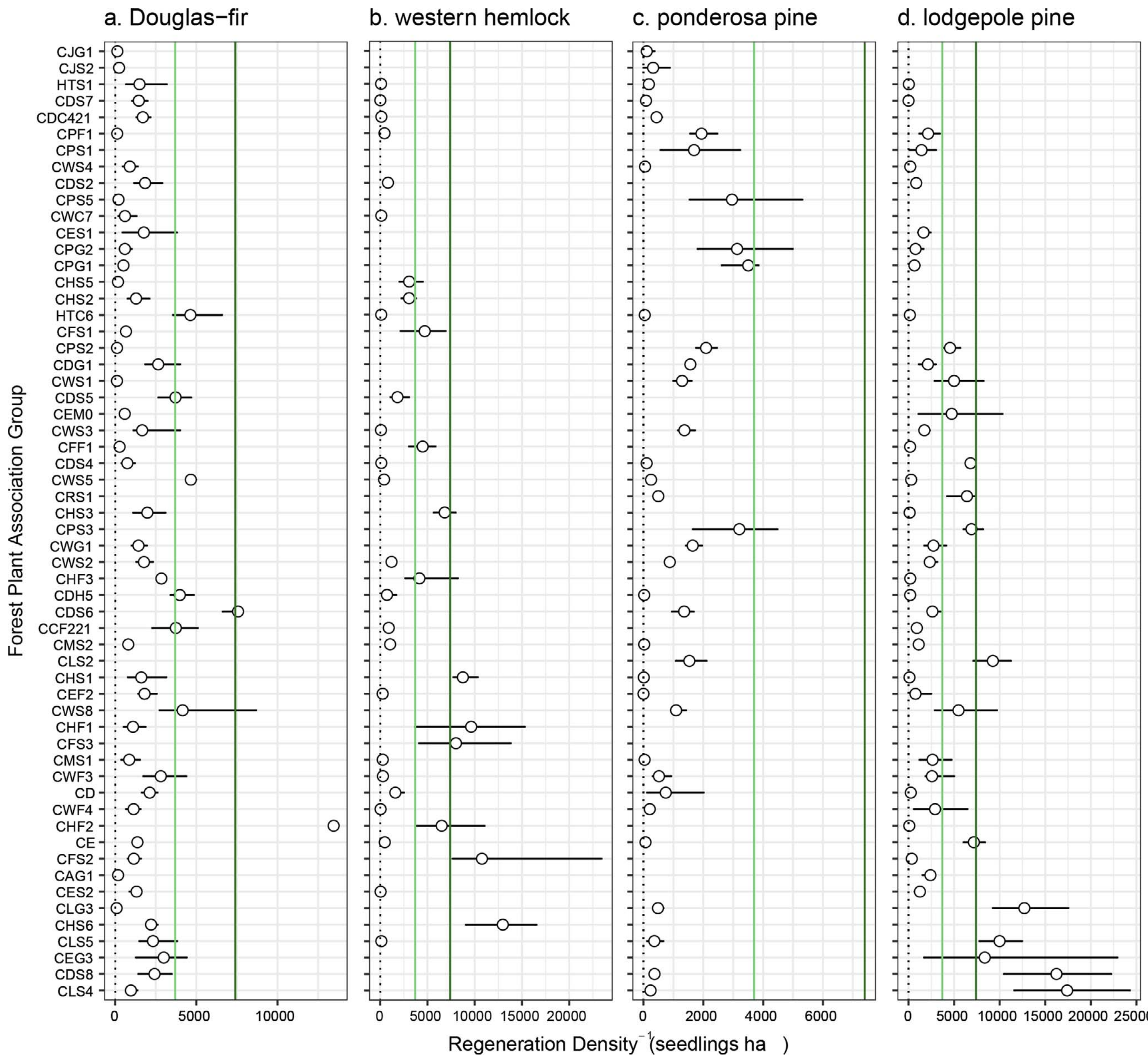

Fig. 5. Ninety-five percent confidence intervals based on 1000 simulations for species-specific root mean-squared error (RMSE) for (a) Douglas-fir, (b) western hemlock, (c) ponderosa pine, and (d) lodgepole pine. RMSE presented for models of Forest Plant Association Group with minimal adequate sample size (n ≥ 120) in Oregon and Washington state in the Pacific Northwest region of the United States and sorted by increasing average total (all species) RMSE.

defined to be at least 120 subplots. This sub-setting resulted in 58 FPAGs for which both model development and evaluation were performed (Table 1). Among these FPAGs the median number of subplots where regeneration was observed was 146, the median number of subplots where no regeneration was observed was 146, and the median number of subplots where more than one species of regeneration was observed was 53. Where regeneration was occurring, the mean and maximum number of species regenerating were 1.5 (SE 0.01) and 8 (Table 1); the mean number of species regenerating drops to 0.7 (SE 0.01) when subplots with no regeneration are included. The mean and maximum number of overstory species were 2.1 (SE 0.07) and 8; none of these statistics were substantially affected by filtering.

Prediction variables obtained from FVS simulations were explored using generalized linear models and correlation analysis. Across the 211 FPAGs, regression analysis showed that many of the predictor variables commonly collected in stand exams generally exhibited weak correlations with regeneration density. Some of the strongest available predictor variables were stand density index (SDI; Reineke, 1933), basal area (BA), trees hectare^{-1}, and FVS computed canopy cover (CANCOV; Crookston and Stage, 1999). Using this information, along with empirical knowledge and an awareness of variates commonly available in FVS, the SDI, BA, and CANCOV variables were selected as predictor variables for estimating regeneration species and density by imputation in all of the 211 FPAG-specific models.

To allow for a randomness component and to preserve naturally occurring species-count combinations, a value of k = 10 was chosen and one of the k = 10 nearest neighbors was randomly selected for imputation. This value of k = 10 allowed for additional flexibility in the models in accordance with the study's objectives to create models that, if a user had a priori knowledge of a target stand and so selected, a user could predict for above or below average regeneration and still have a random component for imputed regeneration values between target stands with similar characteristics. If a user selected for above or below average regeneration, nearest neighbors were sorted based on total seedling stem counts (regardless of species) from highest to lowest, with the top 30% representing above average regeneration and the lowest 30%

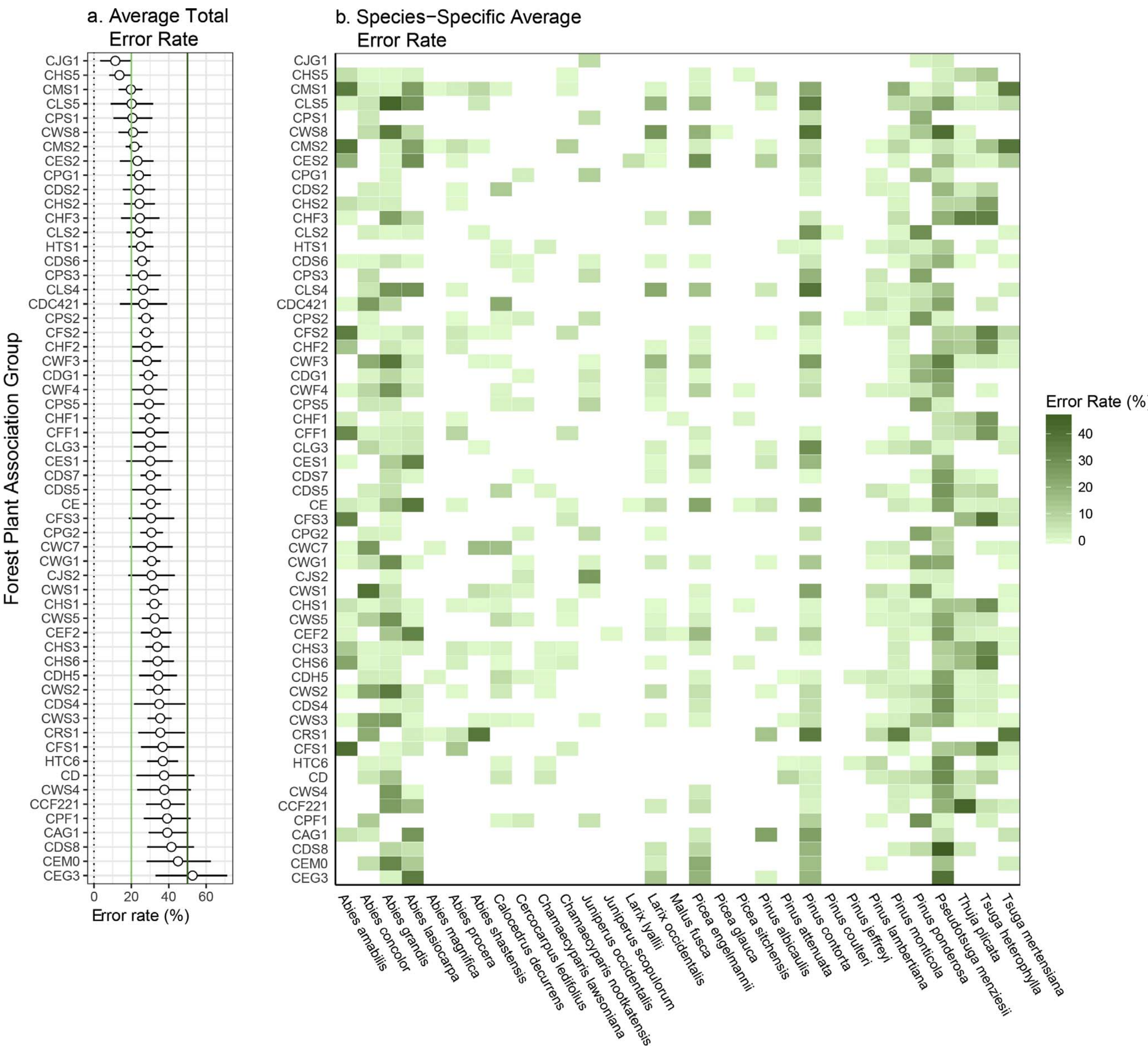

Fig. 6. Ninety-five percent confidence intervals for (a) average Total Error Rate and (b) Average Species-specific Error Rate for Forest Plant Association Group with minimal adequate sample size (n ≥ 120) in Oregon and Washington state in the Pacific Northwest region of the United States and based on 1000 simulations.

representing below average regeneration. Depending on the user's choice, one nearest neighbor was then randomly selected for imputation from the respective above or below pool of $k * 0.3$ nearest neighbors. However when no preference was stated all $k = 10$ nearest neighbors were considered and one randomly selected for imputation. Although only summary statistics for the case where no preference of above or below average regeneration are presented here, these model objective restrictions did influence the choice of k. The k value of 10 was selected to optimize the tradeoff between still having a random component for when above/below average regeneration was specified and maintaining a low enough value of k that it would not negate the nearest neighbor aspect of the model (i.e.: that not all target stands would have the entire reference population available as nearest neighbors).

While both MSN and GNN were considered as possible imputation methods for this study, preliminary comparisons between the methods showed that MSN outperformed GNN for the majority of FPAG models. The poorer performance of GNN might follow from the fact that GNN's use of canonical correspondence analysis assumes environmental factors for ordination and those factors may not be reflected in the prediction variables examined for use in these models (Ohmann and Gregory 2002, Eskelson et al. 2009b). Based on this and coupled with the frequency of MSN over GNN use in regeneration literature, MSN method was used as the method to find nearest neighbors for the models.

Within each FPAG, data splitting was employed and subplots were randomly divided into testing (25%) and training (75%) data sets to test the accuracy of the FPAG-specific models. Models were developed solely with the training data set and the testing data set was set aside for evaluation purposes. Fig. 3 depicts the MSN with $k = 10$ modeling approach used in this study for each FPAG. In brief, all model development and analyses were conducted in R and *yaimpute* (R Development Core Team, 2008, Crookston and Finley, 2008) was used to find nearest neighbors based on a MSN with $k = 10$ imputation approach. After candidate nearest neighbor subplots were selected, one nearest neighbor subplot was selected at random and all of the seedling species and count data was imputed.

Model performance and predictive capability was examined in four different ways:

(i) ability to correctly predict total seedling density (regardless of species);

(ii) ability to correctly predict seedling density for a given species;
(iii) ability to correctly predict the presence or absence of any regeneration; and
(iv) ability to correctly predict seedling species-density combinations.

To more precisely quantify the range in potential values resulting from a single selection of k = 10 nearest neighbors at random, 1000 imputation simulations were performed with the same testing and training data set. The resulting distributions of evaluation statistics were then summarized for the 1000 simulations.

To address methods (i) and (ii) for comparing model performance, evaluation statistics of bias, mean absolute deviation (MAD), and root mean-squared error (RMSE) of regeneration stems per ha^{-1} (from here on abbreviated as TPH) were calculated for each individual simulation as follows:

$$Bias = \sum_{i=1}^{n}\left[\frac{(y_i-\overline{y}_i)}{n}\right]$$

$$MAD = \sum_{i=1}^{n}\left[\frac{|y_i-\overline{y}_i|}{n}\right]$$

$$RMSE = \sqrt{\sum_{i=1}^{n}\left[\frac{(y_i-\overline{y}_i)^2}{n}\right]}$$

where y_i and $\overline{y}_i$ are the observed TPH and predicted TPH, respectively, for each target plot i, and n is the number of target plots in a given FPAG. These evaluation statistics were calculated for Total TPH (regardless of species) and for Species-specific TPH (individual species). Evaluation statistics for species-specific TPH were only calculated for species able to regenerate in a given FPAG. Ability for a species to regenerate in a given FPAG was determined before modeling based on whether that species was recorded as regenerating with that FPAG, and therefore able to appear in the randomly selected training data set as regeneration (i.e., able to be imputed as regeneration).

To address methods (iii) and (iv) for comparing model performance, error rate (ER) in prediction was examined with respect to the total presence or absence of regeneration. To address the method (iii), Total ER was calculated by evaluating ER as the frequency with which a model incorrectly predicted the total presence or absence of regeneration regardless of species. Species-specific ER was used to address method (iv), by evaluating ER as the frequency with which a model incorrectly predicted the presence or absence of regeneration with regard to species. Both statistics were calculated as follows:

$$ER = \frac{FP + FN}{TP + TN + FP + FN}\cdot 100\%$$

where TP is a true positive (present species correctly identified as present), FP is a false positive (absent species incorrectly identified as present), TN is a true negative (absent species correctly identified as absent) and FN is a false negative (present species incorrectly identified as absent).

Although the field collection protocol for the data used in this study makes no allowance for error if the regeneration count for a single species is 5 or less on a microplot, there is a 20% error tolerance if the count is > 5 (USFS 2015). Therefore, FPAG-specific models with low Total ER were defined as models which incorrectly predicted the presence or absence of any regeneration for less than 20% of the subplots within a FPAG; moderate error rates were those with Total ER between 20 and 50%; and high error had Total ER > 50%.

Similarly, models were considered to have low Total RMSE for values under 3700 TPH, moderate Total RMSE values were between 3700 and 7400 TPH, and high Total RMSE were those values > 7400 TPH. Threshold values for low and moderate Total RMSE equate to roughly 5 and 10 seedlings per imputed microplot, respectively. It is important to note that these designations of low, moderate, and high RMSE are with respect to the data underlying these models (seedlings counts recorded on a 2.07 m radius microplot); they do not reflect what might be low, moderate, or high RMSE with respect to implementation in FVS for a particular FPAG. A model was then categorized based on the average RMSE and average Total ER across the 1000 simulations.

3. Results

The following model evaluation and summary statistics apply to the 58 FPAG-specific models examined for model assessment.

3.1. Estimating total seedling density (regardless of species)

Twenty-five models had low to moderate average Total RMSE across the 1000 simulations, with 10 of the 58 models having low average RMSE (Table 2). Additionally, all but one model (CEG3) had an average Total bias under 3700 TPH, and all but three models (CEG3, CLS4, CLS5) had an average MAD within 7400 TPH as well. Although Total RMSE, bias, and MAD did not appear to be affected by sample size, RMSE and MAD tended to increase with increasing interquartile range (IQR) of total seedling counts (regardless of species) (Fig. 4). Variability in bias also generally increased with RSME (Fig. 4c).

The five FPAG-specific models with the lowest average Total RMSE were western juniper/bluebunch wheatgrass (CJG1; 665), western juniper/mountain big sagebrush (CJS2; 1444); tanoak/California huckleberry (HTS1; 2205), Douglas-fir/mallow ninebark (CDS7; 2630), and Douglas-fir-white fir (CDC421; 2988) (Fig. 4a). There were thirty-three models that exhibited high Total RMSE. The FPAG at the lowest end of the high Total RMSE category was Douglas-fir/ceanothus-manzanita-Oregon boxleaf (CDS4; 7441). Other FPAGs with high Total RMSE ranged from western hemlock/bride's bonnet-twinflower (mesic; CHF3; 9602) and western hemlock/sweet after death-threeleaf foamflower (dry; CHF2; 14,124), to subalpine fir-spruce/grass (CEG3; 19,498) and lodgepole pine/grouse whortleberry (CLS4; 25,225).

Confidence intervals based on the 1000 simulations suggest some of the models consistently produced negative biases (overestimation) (Fig. 4c). On average, the most extreme case of consistent overestimation is in the model for Park (CAG1; an FPAG for park-like subalpine-fir, whitebark and lodgepole pine plant associations), which had an average Total bias of −2809 TPH. There were fewer models that tended towards consistent underestimation, with the most extreme case of underestimation being subalpine fir-spruce/grass (CEG3) with an average Total bias of 3967 TPH.

3.2. Estimating seedling density for a given species

Evaluation statistics were calculated for 660 species by FPAG model combinations. Fig. 5 (a–d) shows 95% confidence intervals for Species-specific RMSE for the four most common species in the study area: Douglas-fir, western hemlock, ponderosa pine, and lodgepole pine (for Species-specific MAD and bias results see Appendix Figs. A.1b–e and A.2b–e, respectively).

Of all species by FPAG combinations, the two most extreme values for average Species-specific RMSE were produced by lodgepole pine in the lodgepole pine/grouse whortleberry (CLS4; 17,403) and Douglas-fir/blueberry (CDS8; 16,235) models, which were the two models with the highest average Total RMSE. The longest confidence intervals were produced by lodgepole pine in CEG3 (1687–22,918 TPH) and western hemlock in CFS2 (7652–23,420 TPH). Of the four most common species in the study area, ponderosa pine had lower extreme values of average Species-specific RMSE across FPAG models. As expected, the higher values of average Species-specific RMSE for ponderosa pine were produced by pine-dominated FPAGs (abbreviations beginning with CP).

At the species level, confidence intervals suggest that 128 of the 660 species by FPAG combinations resulted in consistent negative or positive bias, with 49 of the 58 models consistently over or underestimated certain species. The CE model consistently over or underestimated the greatest number of species (6); this is 33% of the 18 species with potential to regenerate in this FPAG. The CEM0 model consistently overestimated the greatest percent of species able to regenerate in that FPAG (4 out of 9 species). No model showed consistent positive or negative bias for over half the species defined as able to regenerate in an FPAG.

The most common consistently overestimated species were lodgepole pine and Douglas-fir, which were each consistently overestimated by 11 models. Only two species had consistent negative bias in fifty percent of the models in which they appeared, however both species only occurred in two models: Subalpine larch (*Larix lyallii*; CE, CES2; CES2) and Oregon crab apple (*Malus fusca*; CEF2, CHF1; CHF1)

3.3. Estimating the presence or absence of any regeneration

Only one model did not have a low to moderate average Total ER, and that was subalpine fir-spruce/grass (CEG3), which on average incorrectly predicted the presence or absence of any regeneration for 52.7% of the target plots in a simulation (Fig. 6a). Most models had a moderate average Total ER, however four models had a low average Total ER. CJG1, which also had the lowest RMSE, had the lowest average Total ER at 11.4%. Total ER did not have a strong relationship with IQR or number of subplots. Total ER did have a quadratic relationship with percentage of subplots with regeneration, as might be expected.

3.4. Estimating seedling species-density combinations

The most substantial average Species-specific ER contributions appear to be primarily from the most common species in the study area, with some of the largest contributions observed in FPAGs that included Douglas-fir, western hemlock, lodgepole pine, grand fir (*Abies grandis*), subalpine fir (*Abies lasiocarpa*), and white fir (*Abies concolor*) (Fig. 6b). The highest average Species-specific ER is 47.0% for Douglas-fir in the Douglas-fir/blueberry (CDS8). The two models with the lowest average across species for average Species-specific ER (average of the average Species-specific ERs across all species regenerating in that FPAG) are Douglas-fir/spiraea-common snowberry-kinnikinnick (CDS6) and tanoak/California huckleberry (HTS1) at 3.1%; both FPAGs had the largest contribution to that average from Douglas-fir.

3.5. Overall model performance (RMSE and Total ER comparison)

Thirty-three models out of the 58 FPAG-specific models developed in this study had high RMSE or high Total ER, with the other 43% of the models falling within the low to moderate RMSE and Total ER categories. Twenty-five models had moderate Total ER and low to moderate RMSE, and one model had both low Total ER and low RMSE: western juniper/bluebunch wheatgrass (CJG1).

4. Discussion

The majority of FPAG-specific models developed in this study exhibited what we classified as low to moderate values of RMSE as well as low to moderate Total ER. These results are comparable to those presented in similar studies using MSN imputation methods for regeneration. Froese et al. (2003) reported high regeneration stems ha^{-1} values for RMSE (5897), bias (−541), and MAD (4399) when applying MSN imputation methods for predicting regeneration in British Columbia. Hassani et al., (2004) used MSN imputation to predict regeneration in British Columbia for 16 height-class by species-group variable combinations, and reported evaluation metrics as averages of RMSE (1381–2093 TPH), bias (−143 to 154 TPH), and MAD (576–792 TPH) over these 16 regeneration variables. In this study, the majority of FPAG models for which evaluation statistics were computed fell within the poor performance statistics reported by Hassani et al. (2004) if averaged Total RMSE was considered (see Fig. 4). However if averaged Species-specific RMSE is considered, which might be a closer comparison to the statistics reported by Hassani et al. (2004), 68% of the 660 species by FPAG model combinations resulted in what Hassani et al. categorized as low RMSE.

It is important to note once more that tree species that commonly use sprouting as a means of reproduction were not included in this modeling effort. Therefore values of total TPH predicted by these models do include the possible contribution that sprouting species would make to overall regeneration densities. Along that vein, the three best preforming FPAG models with regard to RMSE are likely relics of their stand structure; meaning that for the species for which regeneration was modeled, recorded regeneration stems ha^{-1} was relatively low in comparison to many of the other 58 FPAGs for which model summary statistics were calculated. FPAGs with higher IQR for total seedling counts, or a greater number of species that could regenerate tended to have less superior model performance. This makes sense as some of these more complex FPAGs tend to have more complex overstories, than for example, CJG1. However, predicting regeneration from an observed subplot using imputation allowed for the inherent variability of these FPAGs to be preserved.

As stated earlier, it is important to note that reported errors are with respect to the observed data used to develop our models and do not necessarily reflect what might be considered as "suitable" errors when used in conjunction with a growth-and-yield model such as FVS. Furthermore, the ingrowth or recruitment (*sensu* Weiskittel et al., 2011) process typically includes additional predictions of small tree growth and mortality, neither of which were examined in this study but both are known to be important factors when modeling the regeneration process. Considerations should be given to the magnitude of our reported errors and implementation framework when using our regeneration models in conjunction with growth models such as FVS. For example, we did not provide predictions of small tree growth and mortality, thus these processes would also need to be incorporated by the user in order to have regeneration, and thus recruitment, function properly over time and represent how these processes may interact.

Despite the requirement of the models to predict based on only input from data collected in common stand exams, the FPAG-specific models performed relatively well. However, model results may be improved by including variables that have stronger predictable relationships to regeneration. The lower than desired correlations between the regeneration and the predictor variables in this study is a similar finding to that reported by Hassani et al. (2004), who speculated that this may indicate that either predictor variables were not useful or that the regeneration is highly variable and thus required more data to capture. The latter is also a possibility for data used in this study, as there were substantially more strata (211 individual FPAG models with three stand variables versus one model with ten overstory-site variables in Hassani et al.) and the variables were *a prior* limited to those available from common stand exams and FVS simulations.

Contrary to these limitations, a primary benefit to imputation models is that the incorporation of additional samples or predictor variables as they become available is relatively simple. The addition of more predictor variables and/or samples would likely increase the range of observations, allowing for more detailed nearest neighbor selection, using variables such as seedling height if available (Froese et al. (2003), Hassani et al., 2004) to further stratify the data.

4.1. Future directions

Utilizing the freely available FIA and PNW data from Oregon and Washington presented an unparalleled opportunity in terms of sample size and inference space. However further consideration could be given to obtaining information for the very wide range of disturbance conditions that may influence natural regeneration, including those that contribute to unusual or unique stand conditions. These considerations could help guarantee that the reference data reflect the variability expected and potentially observed in the population. Additional considerations when using FIA and NFS monitoring data are the sample design and the intentions or goals of the sample, particularly the scale, sample frequency and detection issues. FIA plots generally meet program quality standards with respect to disturbance measures and thus consequential regeneration measurements; however this is largely due to the fact that disturbance tends to be rarely observed (Pollard et al., 2006). Schroeder et al. (2014) suggest that overall improvements in spatial and temporal sampling of disturbances may be improved by as much as 28% by combining observations from Landsat time series with FIA data.

5. Conclusions

As compared to other modeling techniques, such as parametric regression analysis, imputation approaches do not require distributional assumptions, and are inherently multivariate in nature. For regeneration prediction, the use of imputation has the distinct advantage of predicting regeneration densities by species for a large range of plant associations; utilizing existing inventory and permanent plot data; and requiring minimal stand information (density and canopy cover) to return reasonable predictions for many stands. Furthermore, these models can be easily updated as new data become available and the method outlined here is easily adaptable to other forest systems where FIA data are available.

Prior to this effort, user specification of regeneration densities and species was the only approach available to managers using FVS in the Pacific Northwest. This study is a large first step towards developing a flexible, expandable, and adaptable regeneration models that can be easily incorporated into FVS. Additional steps can be taken to improve individual FPAG-specific models, including a refinement of the FPAG classifications themselves. Potential weighting of different variables or the inclusion of some variables not presently in the models but known to be of importance (e.g., surface fuel loading, shrub density, or topographic position) could lead to improvement in all or a subset of the models.

Acknowledgements

This research was supported by funding from Pacific Northwest Regional Office (Region 6) of the USDA Forest Service and by the ARCS Foundation Scholar program. We would like to particularly thank Tom DeMeo, Regional Ecologist of the Pacific Northwest Region, for his input on the grouping of plant associations into FPAGs. We would also like to thank two anonymous reviewers for their insightful comments. Northern Arizona University and Oregon State University are equal opportunity providers.

Appendix A

See Figs. A.1–A.5.

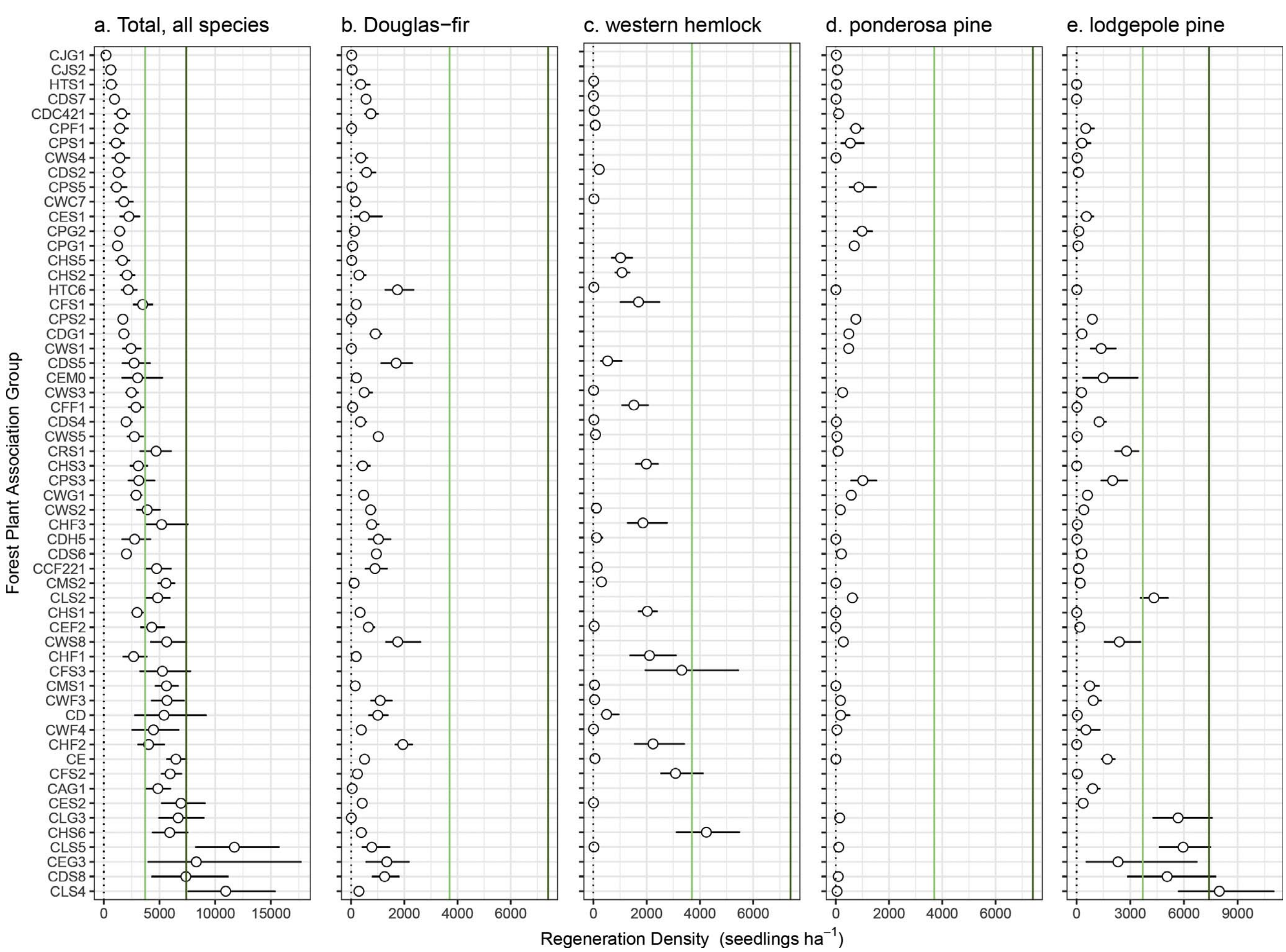

Fig. A.1. Ninety-five percent confidence intervals based on 1000 simulations for (a) average total mean absolute deviation (MAD) (all species), and average species-specific MAD for (b) Douglas-fir, (c) western hemlock, (d) ponderosa pine, and (e) lodgepole pine. MAD presented for models of Forest Plant Association Group with minimal adequate sample size (n ≥ 120) in Oregon and Washington state in the Pacific Northwest region of the United States.

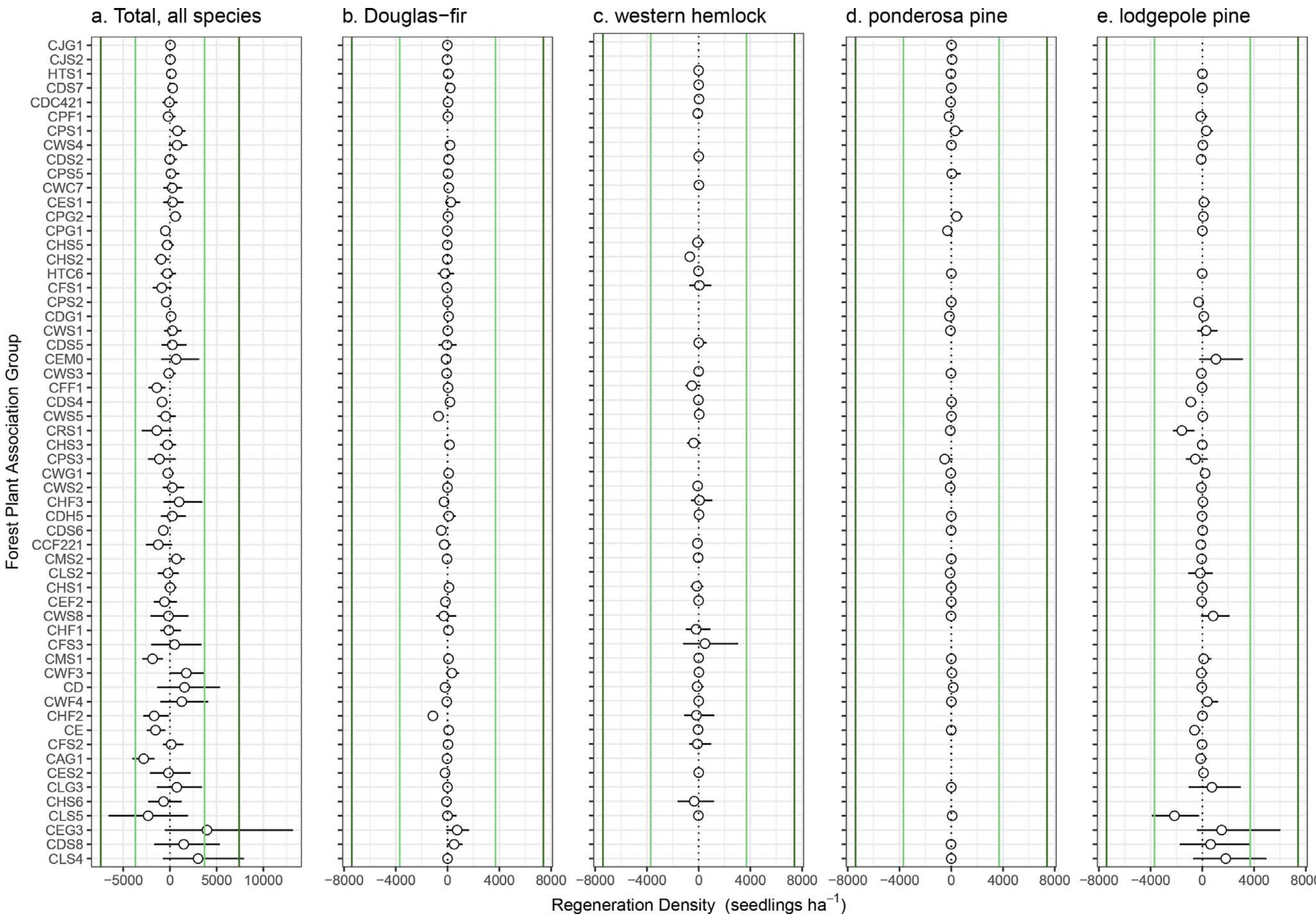

Fig. A.2. Ninety-five percent confidence intervals based on 1000 simulations for (a) average total bias (all species), and average species-specific bias for (b) Douglas-fir, (c) western hemlock, (d) ponderosa pine, and (e) lodgepole pine. Bias summaries presented for models of Forest Plant Association Group with minimal adequate sample size (n ≥ 120) in Oregon and Washington state in the Pacific Northwest region of the United States.

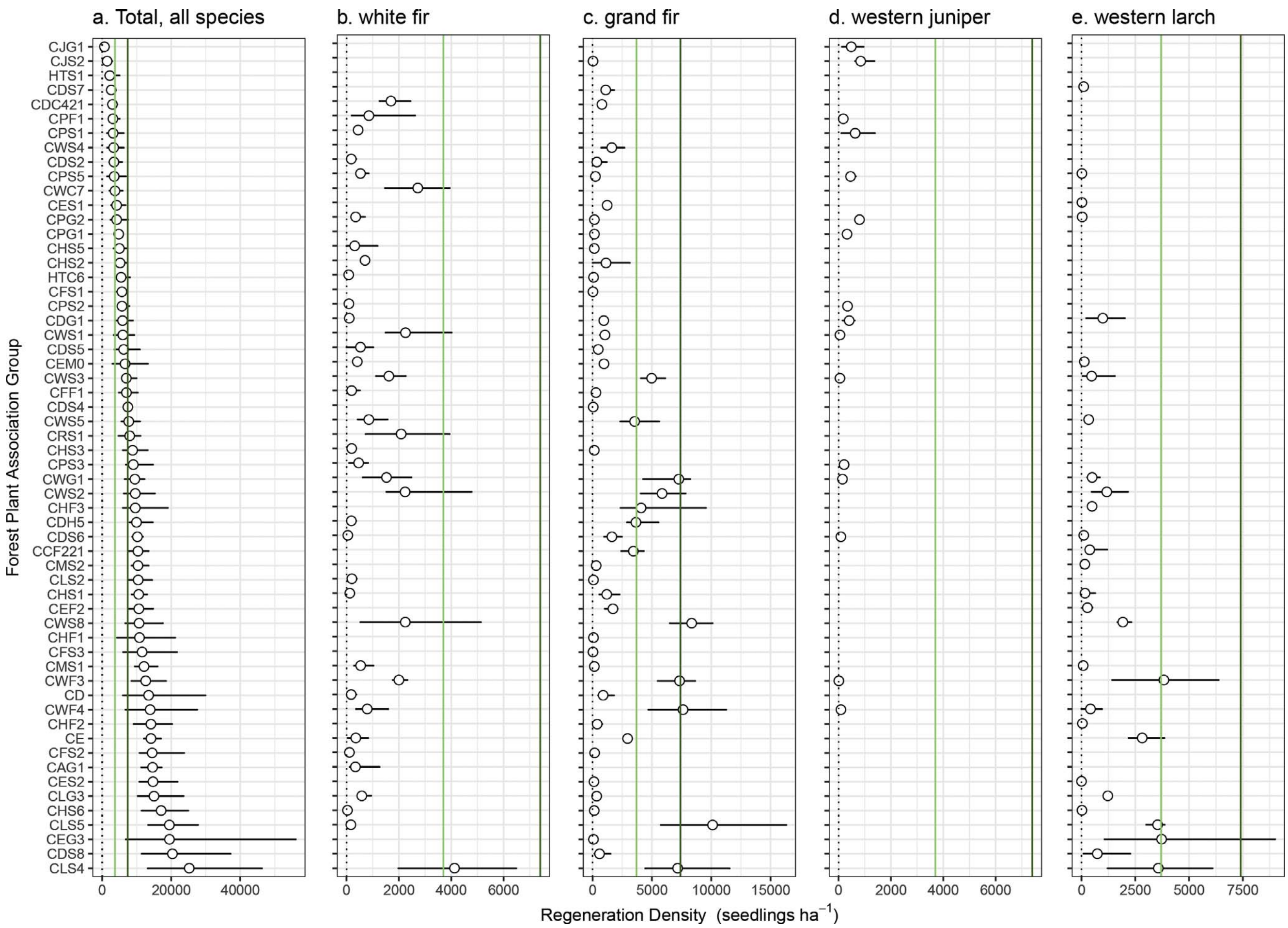

Fig. A.3. Ninety-five percent confidence intervals based on 1000 simulations for (a) average total root mean-squared error (RMSE) (all species), and average species-specific RMSE for (b) white fir, (c) grand fir, (d) western juniper, and (e) western larch. RMSE presented for models of Forest Plant Association Group with minimal adequate sample size (n ≥ 120) in Oregon and Washington state in the Pacific Northwest region of the United States.

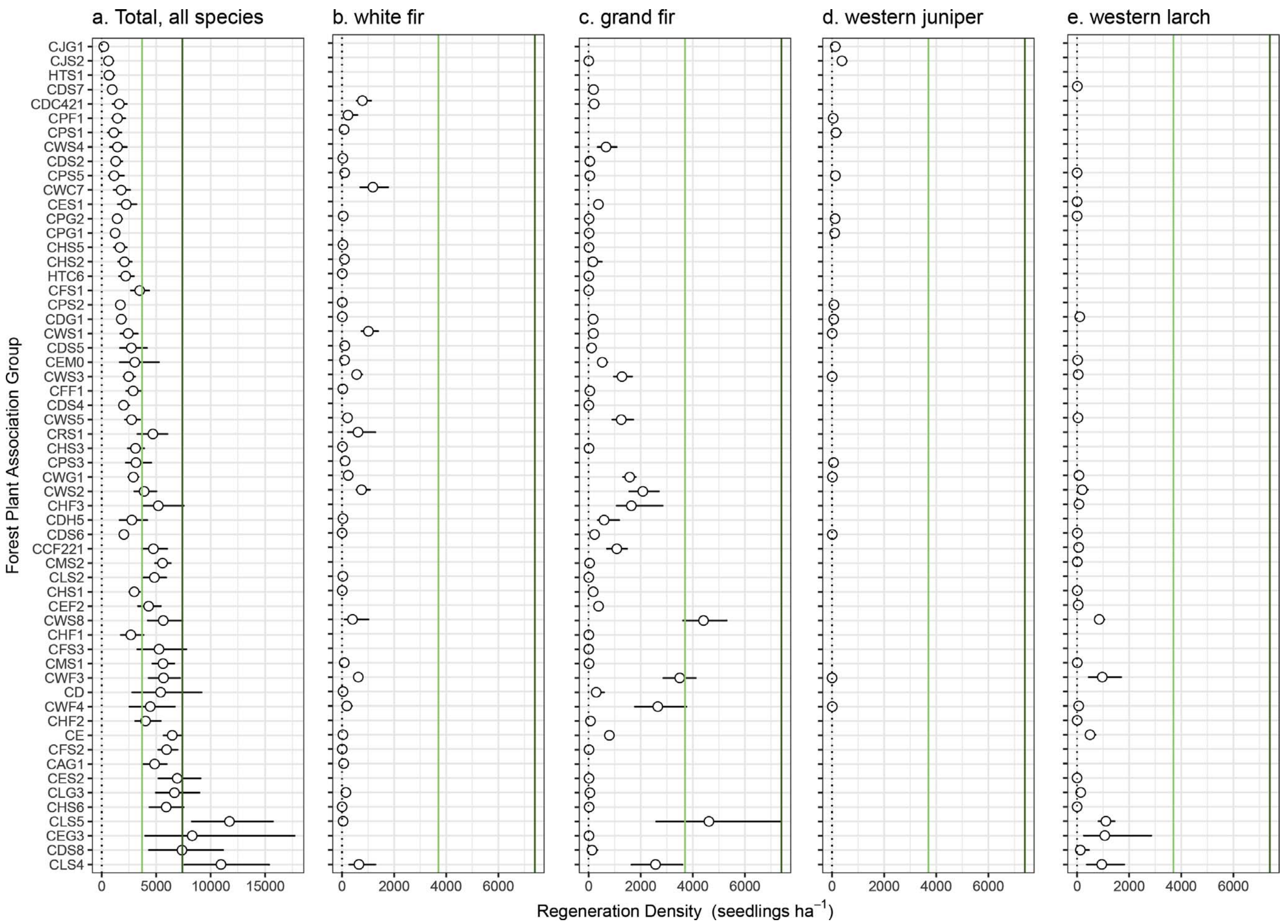

Fig. A.4. Ninety-five percent confidence intervals based on 1000 simulations for (a) average total mean absolute deviation (MAD) (all species), and average species-specific MAD for (b) white fir, (c) grand fir, (d) western juniper, and (e) western larch. MAD presented for models of Forest Plant Association Group with minimal adequate sample size (n ≥ 120) in Oregon and Washington state in the Pacific Northwest region of the United States.

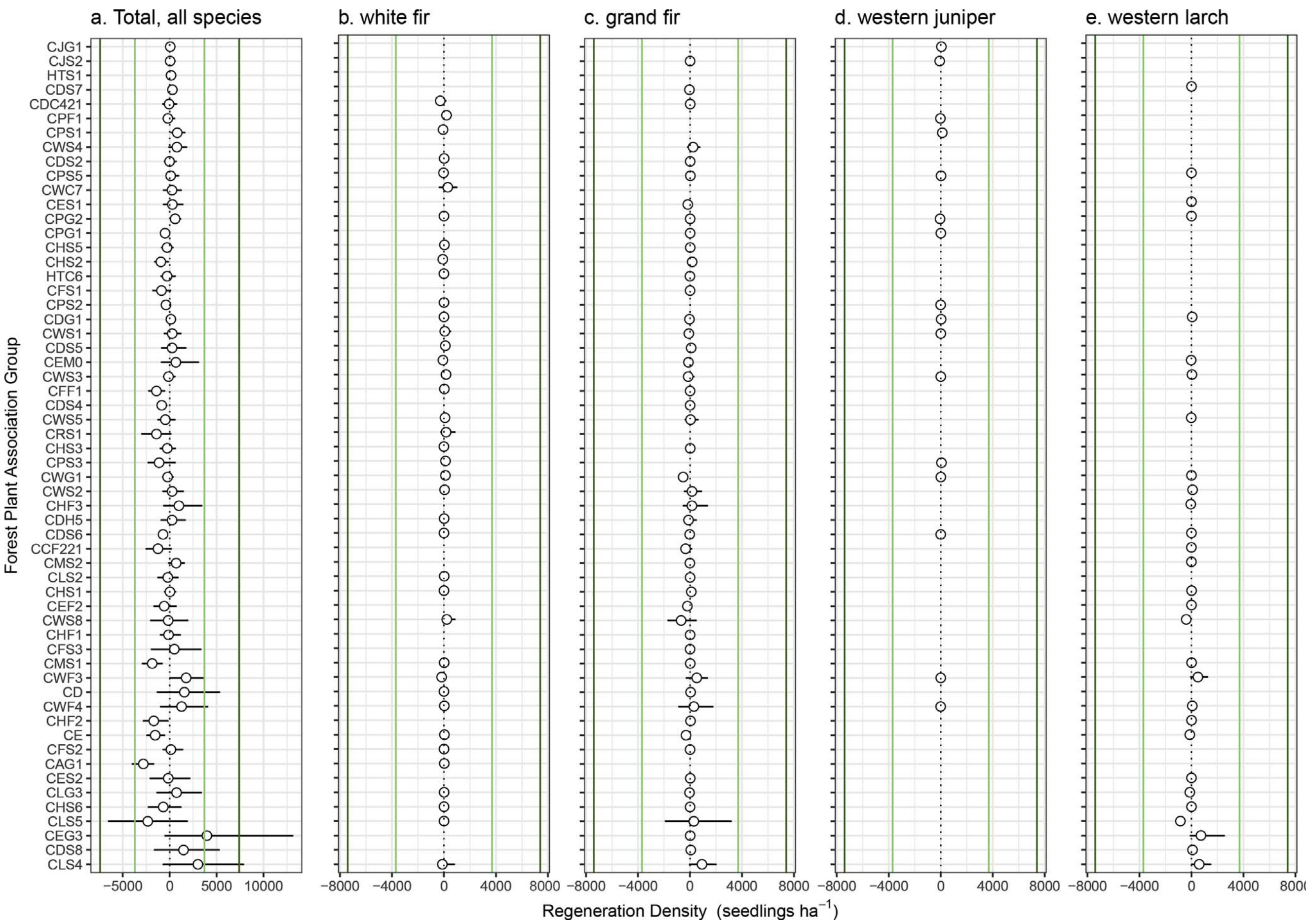

Fig. A.5. Ninety-five percent confidence intervals based on 1000 simulations for (a) average total bias (all species), and average species-specific bias for (b) white fir, (c) grand fir, (d) western juniper, and (e) western larch. Bias presented for models of Forest Plant Association Group with minimal adequate sample size (n ≥ 120) in Oregon and Washington state in the Pacific Northwest region of the United States.

References

Bechtold, W.A., Scott, C.T. 2005. The forest inventory and analysis plot design. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station, p. 37–52.

Crookston, N.L., Finley, A.O., 2008. yaImpute: An r package for knn imputation. J. Stat. Softw. 23 (10), 1–16.

Crookston, N.L., Stage, A.R. 1999. Percent canopy cover and stand structure statistics from the Forest Vegetation Simulator. US For. Serv. Gen. Tech. Rep. RMRS-GTR-24.

Crotteau, J.S., Ritchie, M.W., Varner, J.M., 2014. A mixed-effects heterogeneous negative binomial model for postfire conifer regeneration in Northeastern California USA. For. Sci. 60 (2), 275–287.

Dixon, G.E. 2013. Essential FVS: A user's guide to the Forest Vegetation Simulator. U.S. Department of Agriculture, Forest Service, Forest Management Service Center, Fort Collins, CO. 266 p.

Ek, A.R., Robinson, A.P., Radtke, P.J., Walters, D.K., 1997. Development and testing of regeneration imputation models for forests in Minnesota. For. Ecol. Manage. 94 (1), 129–140.

Eskelson, B.N., Barrett, T.M., Temesgen, H., 2009a. Imputing mean annual change to estimate current forest attributes. Silva Fennica. 43 (4), 649–658.

Eskelson, B.N.I., Temesgen, H., LeMay, V., Barrett, T.M., Crookston, N.L., Hudak, A.T., 2009b. The roles of nearest neighbor methods in imputing missing data in forest inventory and monitoring databases. Scan. J. For. Res. 24, 235–246.

Ferguson, D.E., Carlson, C.E. 1993. Predicting regeneration establishment with the Prognosis Model. Res. Paper INT-467. U.S. Department of Agriculture, Forest Service, Intermountain Research Station, Ogden, UT, 54 pp.

Ferguson, D.E., Johnson, R.R. 1988. Developing variants for the regeneration establishment model. In: Ek, A.R., Shifley, S.R., Burk, T.E. (Eds.), Proceedings: Forest Growth Modeling and Prediction, August 23–27, 1987, Minneapolis, MN. Gen. Gech. Rep. NC-120. U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station, St. Paul, MN. pp. 369–376.

Ferguson, D.E., Stage, A.R., Boyd Jr., R.J., 1986. Predicting regeneration in the grand fir-cedar-hemlock ecosystem of the northern Rocky Mountains. For. Sci. Monogr. 26, 41.

Fortin, M., DeBlois, J., 2007. Modeling tree recruitment with zero-inflated models: the example of hardwood stands in southern Qúebec Canada. For. Sci. 53, 529–539.

Froese, K.L., LeMay, V., Marshall, P., Zumrawi, A.-A. 2003. Regeneration imputation models for PrognosisBC: IDFdm2 subzone variant, Invermere Forest District. Report prepared for the Forestry Innovation Investment. Ref. No. R02-07, 67 pp.

Hall, F.C. 1998. Pacific Northwest ecoclass codes for seral and potential natural communities. Gen. Tech. Rep. PNW-GTR-418. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 290 p. In cooperation with: Pacific Northwest Region.

Hassani, B.T., LeMay, V., Marshall, P.L., Temesgen, H., Zumrawi, A.A., 2004. Regeneration imputation models for complex stands of southeastern British Columbia. For. Chron. 80 (2), 271–278.

Lacerte, V., Larocque, G.R., Woods, M., Parton, W.J., Penner, M., 2006. Calibration of the forest vegetation simulator (FVS) model for the main forest species of Ontario Canada. Ecol. Model. 199 (3), 336–349.

LeMay, V., Temesgen, H., 2005. Comparison of nearest neighbor methods for estimating basal area and stems per hectare using aerial auxiliary variables. For. Sci. 51 (2), 109–119.

Li, R., Weiskittel, A.R., Kershaw, J.A., 2011. Modeling annualized occurrence, frequency, and composition of ingrowth using mixed-effects zero-inflated models and permanent plots in the Acadian Forest Region of North America. Can. J. For. Res. 41, 2077–2089.

Lillybridge, T.R., Kovalchik, B.L., Williams, C.K., and Smith, B.G. 1995. Field guide for forested plant associations of the Wenatchee National Forest. Gen. Tech. Rep. PNW-GTR-359. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 335 p.

McRoberts, R.E., 2001. Imputation and model-based updating techniques for annual forest inventories. For. Sci. 47 (3), 322–330.

Moeur, M., Stage, A.R., 1995. Most similar neighbor: an improved sampling inference procedure for natural resource planning. For. Sci. 41 (2), 337–359.

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 USA. Can. J. For. Res. 32, 725–741.

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. J. Veg. Sci. 22 (4), 660–676.

Oliver, C.D., Larson, B.C., 1996. Forest stand dynamics. John Wiley & Sons Inc, New York, pp. 520.

Pollard, J.E., Westfall, J.A., Patterson, P.L., Gartner, D.L., Hansen, M., Kuegler, O. 2006. Forest inventory and analysis national data quality assessment report for 2000 to 2003. In: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO, 43 pp.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reineke, L.H., 1933. Perfecting a stand-density index for even-aged forests. J. Ag. Res. 46, 627–638.

Rubin, D.B., 1987. Multiple imputations for non-response in surveys. John Wiley & Sons, New York, United States of America, pp. 258p.

Schroeder, T.A., Healey, S.P., Moisen, G.G., Frescino, T.S., Cohen, W.B., Huang, C., Kennedy, R.E., Yang, Z., 2014. Improving estimates of forest disturbance by combining observations from Landsat time series with U.S. Forest Service Forest Inventory and Analysis data. Remote Sens. Environ. 154, 61–73.

USDA Forest Service (USFS), 2015. FSVeg Common Stand Exam User Guide. Version 2.12. 6. USDA For. Serv. Available: http://www.fs.fed.us/nrm/fsveg/.

USDA Forest Service (USFS), 2014. Land areas of the National Forest Systems, as of September 30, 2013. 2014. USDA For. Serv. FS-383. Available: http://www.fs.fed.us/land/staff/lar/LAR2013/lar2013index.html. (Accessed July 6, 2015).

USDA Forest Service (USFS), 2013. Forest inventory and analysis national program - data and tools - FIA data mart, FIADB Version 5.1. USDA Forest Service. Available: http://apps.fs.fed.us/fiadb-downloads/datamart.html. (Accessed 27 October 2014).

Weiskittel, A.R., Hann, D.W., Kerhsaw Jr., J.A., and Vanclay, J.K., 2011. Forest Growth and Yield Modeling. Wiley. 424 p.