

Tree-level imputation techniques to estimate current plot-level attributes in the Pacific Northwest using paneled inventory data

Bianca N.I. Eskelson¹, Hailemariam Temesgen², Tara M. Barrett³

Abstract: *The Forest Inventory and Analysis program (FIA) of the US Forest Service conducts a nationwide annual inventory. One panel (20% or 10% of all plots in the eastern and western United States, respectively) is measured each year. The precision of the estimates for any given year from one panel is low, and the moving average (MA), which is considered to be the default estimator, can result in biased estimates of current conditions. An alternative to the MA is sought, and studies comparing different alternatives to the MA approach for estimating current forest attributes in the Pacific Northwest are lacking. Paneled data from national forests in Oregon and Washington were used to explore nearest neighbor (NN) imputation methods to project all panels to a common point in time. When using the most recent ground measurements of the panels measured in prior years as ancillary data, tree-level NN imputation outperformed the MA estimator in estimating basal area/ha, stems/ha, volume/ha, and biomass/ha in terms of bias and root mean square error (RMSE) and plot-level NN imputation in terms of RMSE. When basal area/ha, stems/ha, volume/ha, and biomass/ha were summarized by three species groups, tree-level NN imputation outperformed plot-level NN imputation in terms of both bias and RMSE. Tree-level NN imputation outperformed the MA in terms of bias and RMSE for estimating basal area/ha, stems/ha, volume/ha, and biomass/ha for species group ‘pine’ and provided comparable results in terms of bias and RMSE for species groups ‘Douglas-fir’ and ‘other.’*

Keywords: moving average, nearest neighbor imputation, panel, plot-level, tree-level

Introduction

Information on current forest condition is essential to assess and characterize resources and to support management and policy decisions. The 1998 Farm Bill mandates the US Forest Service to conduct annual inventories to provide annual updates of each state’s forest. Only 10% or 20% of all plots in the western and eastern United States, respectively, are measured annually. Because only a portion

¹ Oregon State University; Department of Forest Engineering, Resources and Management; 204 Peavy Hall; Corvallis, OR 97331-5706 USA; e-mail: bianca.eskelson@oregonstate.edu.

² Oregon State University; Department of Forest Engineering, Resources and Management; 204 Peavy Hall; Corvallis, OR 97331-5706 USA.

³ United States Forest Service; Pacific Northwest Research Station; Forest Sciences Lab; 3301 "C" Street, Suite 200, Anchorage, AK 99503-3954 USA.

of the full sample is measured annually, the precision of the estimates for any given year is low. To achieve higher precision, the Forest Inventory and Analysis program (FIA) uses a moving average (MA) as default estimator which combines the data of multiple panels. In the presence of trend, biased estimates will result, if the MA is applied to the end of the period to estimate current conditions. Other approaches to combine data from all panels include: 1) updating unmeasured panel data to the current year with growth models (Lessard et al. 2001); 2) using time series models (Johnson et al. 2003); 3) mixed estimation (Van Deusen 1996); or 4) filling in missing panel data using tree- and plot-level imputation techniques (Gartner and Reams 2001, 2002, McRoberts 2001). Since spatial, temporal, and forest characteristics differ within and among regions it is unclear if any single technique will provide satisfactory results for all regions (Patterson and Reams 2005). It may be necessary to evaluate different methods for a variety of issues and regions. Studies comparing different alternatives to the MA approach for estimating current forest attributes in the Pacific Northwest (PNW) are lacking.

Nearest Neighbor (NN) imputation methods are donor-based, which means that the imputed value was either observed for another unit or was calculated as the average of values from more than one unit. NN imputation can be performed on different levels. Eskelson et al. (2009) have shown that plot-level imputation, that is plot-level attributes (e.g., basal area/ha) are imputed, can provide more accurate results than the MA approach. They found the randomForest (RF) imputation method (Crookston and Finley 2008), which is an extension of classification and regression tree (CART) methods (Breiman 2001), to outperform other NN imputation methods. Imputation can also be performed at the tree-level, that is tree-level attributes (e.g., diameter at breast height (DBH in cm)) are imputed, and the results of the tree-level imputation are then summarized for each plot (e.g., imputed DBH is used to calculate basal area/ha).

Depending on the intended use, tree- and plot-level imputation techniques differ in their predictive abilities and suitability (Gartner and Reams 2002). Plot-level and tree-level NN imputation techniques might have a similar relationship to each other as whole stand growth models, which might not apply in heterogeneous conditions (Curtis and Hyink 1985), have with single-tree growth models, which can provide more detailed information about stand dynamics and structure (Burkhart 1992). Tree-level nearest neighbor (NN) imputation techniques have been successfully used to estimate tree volumes and heights (Korhonen and Kangas 1997), single-tree biomass (Fehrmann et al. 2008) as well as 5-year diameter growth and bark thickness (Sironen et al. 2001, 2003, 2008).

The objectives of this study are to: 1) use paneled data from the PNW to estimate current forest attributes (see Table 1) using tree-level imputation methods and compare their performance against the MA and the estimates based only on the data from the current panel; 2) examine the performance of tree-level imputation methods to estimate current forest attributes by species groups; and 3) compare tree-level and plot-level imputation results.

Methods

Data

The data used in this study consist of 618 plots from six national forests that were collected as part of the Pacific Northwest Region's Current Vegetation Survey (CVS) of the US Forest Service. The plots were installed between 1993 and 1997 and remeasured in 2000. The particular national forests sampled were the Colville (28), Mt. Hood (111), Ochoco (82), Rogue River (70), Wallowa-Whitman (199), and Winema (128).

Five plots are installed in each basic CVS sampling unit, which is one hectare (ha) in size. Each plot consists of three permanent circular, nested subplots of different sizes in which trees are measured depending upon their DBH. For a detailed description of the CVS inventory see Max et al. (1996). Tree height (HT in m) is only subsampled and missing HTs were filled using height models developed in Barrett (2006) for live trees with DBH of 12.7 cm or larger. Volume and biomass equations from the US Forest Service were used to calculate gross cubic-meter volume and total gross oven dry weight biomass (USDA 2000). For each plot, basal area in m² per ha (BA), stems per ha (SPH), volume in m³ per ha (VOL), and biomass in tons per ha (BIOT) were calculated and summarized (Table 1). BA, SPH, VOL, and BIOT were also calculated for each of the following three species groups: 1) 'Douglas-fir'; 2) 'pine' including all occurring pine species; and 3) 'other' including other conifers and hardwoods. Basal area in larger trees (BAL in m²) was calculated for each tree. Ingrowth for each plot was determined by calculating BA, SPH, VOL, and BIOT for all trees that were present in 2000 but not present at the first measurement occasion. BA and SPH were also calculated for small trees with DBH larger than 2.54 cm and smaller than 12.7 cm.

Table 1: Summary of plot-level variables in 2000.

Variable	Minimum	Mean	Maximum	Std
Basal area (m ² /ha)	0.24	24.32	105.35	19.00
SPH (stems/ha)	1	305	1517	221
Volume (m ³ /ha)	0.66	224.82	1444.74	221.04
Biomass (tons/ha)	0.58	134.09	800.64	132.64
Canopy cover (%)	0	54	97	29
Slope (%)	0	23	83	17
Elevation (m)	274	1389	2377	321
Annual precipitation (In cm) (scaled * 100)	577	683	817	48
Annual mean temperature (°C) (scaled * 100)	60	579	1067	166

The data set comprises 30,709 trees in 33 species. The most common species in decreasing order are Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco), ponderosa pine (*Pinus ponderosa* C. Lawson), grand fir (*Abies grandis* (Douglas ex D. Don) Lindl.), lodgepole pine (*Pinus contorta* Douglas ex Louden), white fir

(*Abies concolor* (Gord. & Glend.) Lindl. ex Hildebr.), and western hemlock (*Tsuga heterophylla* (Raf.) Sarg.) (see Eskelson et al. 2009 for details).

In NN imputation methods, ancillary variables are those variables that are measured on all units. Thematic Mapper (TM) images from 2000 were extracted from the national land-cover database 2001 [NLCD 2001 (Homer et al. 2004)] and used as ancillary data. The raw imagery bands 1 to 5 and band 7 (TM1, TM2, TM3, TM4, TM5, TM7) as well as the Tasseled Cap (TC) transformations (TC1 – TC6) (Kauth and Thomas 1976) were used. The normalized difference vegetation index (NDVI) and three commonly used band ratios (band 4 to band 3 (R43), band 5 to band 4 (R54), and band 5 to band 7 (R57)) were calculated. Percent canopy cover was extracted from the NLCD 2001 (Homer et al. 2004).

The following climate and topography variables for plot locations were additionally used as ancillary data: Annual precipitation and mean annual temperature (Table 1) [Data source: DAYMET Daily Surface Weather Data and Climatological Summaries (Thornton et al. 1997, Thornton and Running 1999)], elevation (EL in m) and transformations (EL², ln(EL)) [Data source: CVS inventory], and slope (%) and aspect (degrees) and transformations (cosine(aspect), sine(aspect), cosine(aspect)*slope, and sine(aspect)*slope) [Data source: 30 m digital elevation model using Arc Workstation GRID surface functions and commands (Environmental Systems Research Institute 1991)]. These climate, topography, and satellite variables have been successfully used as ancillary data for NN imputation methods in previous studies (e.g., Eskelson et al. 2009, Ohmann and Gregory 2002).

Imputation techniques

Panel data is a special case of inventory data with measurements taken at different times. All plots were remeasured in 2000. In order to mimic a panel system with the available data 25% of the plots (154) were randomly assigned to *P4* and the remaining 75% of the plots (464) were assigned to *P1*,

P2, and *P3* based on their year of installation. This resulted in *P1*, *P2*, and *P3* having different sizes for each iteration (Table 2).

Table 2: Number of plots measured by year of installation and corresponding panel assignment. All plots listed were remeasured in 2000.

Year of Installation	# of Plots	Assigned Panel
1993	7	1
1994	229	1
1995	223	2
1996	158	3
1997	1	3

The variables of interest (Y) in this study were BA, SPH, VOL, and BIOT. Their observed mean value in the year 2000 was calculated as:

$$\bar{Y}_{OBS} = \sum_{i=1}^n Y_i / n \quad [1]$$

where Y_i is the observed Y value of the i^{th} plot in 2000 and $n = 618$. The observed mean value was used as best available estimate of the true mean.

For each plot in $P4$, BA, SPH, VOL, and BIOT were calculated using the tree data from $P4$. The mean values of Y for the year 2000 (SAMPLE25 estimator) were calculated as:

$$\bar{Y}_{SAMPLE25} = \sum_{i:Y_i \in P4} Y_i / n_4 \quad [2]$$

where Y_i is the observed Y value of the i^{th} plot, and n_4 is the number of plots in $P4$.

The MA estimator, the FIA default method is:

$$\bar{Y}_{MA(4)} = 0.25 * \bar{Y}_{t-3,i} + 0.25 * \bar{Y}_{t-2,i} + 0.25 * \bar{Y}_{t-1,i} + 0.25 * \bar{Y}_{t,i} \quad [3]$$

where $\bar{Y}_{t-3,i}$, $\bar{Y}_{t-2,i}$, $\bar{Y}_{t-1,i}$, and $\bar{Y}_{t,i}$ are the mean values of the variables of interest of $P1$, $P2$, $P3$, and $P4$, respectively. The MA takes into account that the panels include different numbers of plots. Instead of equal weighting of the panels a weighted version of [3] is proposed:

$$\bar{Y}_{WMA(4)} = w_{t-3} * \bar{Y}_{t-3,i} + w_{t-2} * \bar{Y}_{t-2,i} + w_{t-1} * \bar{Y}_{t-1,i} + w_t * \bar{Y}_{t,i} \quad [4]$$

where w_{t-3} , w_{t-2} , w_{t-1} , and w_t are the weights of $P1$, $P2$, $P3$, and $P4$, respectively. Larger weights were chosen for $P3$ and $P4$ ($w_{t-1} = w_t = 0.3$) than for $P1$ and $P2$ ($w_{t-3} = w_{t-2} = 0.2$). MA(4) and WMA(4) will be referred to as MA and WMA, respectively.

Instead of using the previous measurements to fill in the Y values for $P1$, $P2$, and $P3$, as is done with MA and WMA, the current Y values of $P1$, $P2$, and $P3$ were imputed using tree-level RF and plot-level RF imputation. Target data are units that have ancillary variables measured only (e.g., trees or plots in $P1$, $P2$, and $P3$). Reference data are units where both variables of interest and ancillary variables were measured (e.g., trees or plots in $P4$). RF imputation was employed using the yaImpute R package (Crookston and Finley 2008). Details on RF imputation can, for example, be found in Hudak et al. (2008).

For tree-level RF, the target trees were assumed to be non-sampled trees lacking inventory data in 2000. DBH, HT, and mortality for each target tree were imputed using DBH, HT, and BAL at the previous measurement (DBHocc1, HTocc1, and BALocc1) as ancillary data. The reference trees constituted the pool of potential trees with inventory and ancillary data ($P4$), which could be selected to impute the DBH, HT, and mortality for the target trees. Ingrowth of BA, SPH, VOL, and BIOT was imputed at the plot-level using BA and SPH of small trees at

the previous measurement as well as the available climate, topography, and satellite data as ancillary data. BA, SPH, VOL, and BIOT were calculated for each plot based on the imputed tree data and the imputed ingrowth.

For plot-level RF the previous measurements of the four variables of interest (BAocc1, SPHocc1, VOLocc1, BIOTocc1) were used as ancillary data since this was found to provide better imputation results than using climate, topography, and satellite data in a previous study. For more details see Eskelson et al. (2009).

For both tree-level and plot-level RF the overall mean of the variables of interest for the year 2000 was estimated as:

$$\bar{Y}_{IMP} = \left[\sum_{i:Y_i \in P1} Y_{imp,i} + \sum_{i:Y_i \in P2} Y_{imp,i} + \sum_{i:Y_i \in P3} Y_{imp,i} + \sum_{i:Y_i \in P4} Y_{t,i} \right] / n \quad [5]$$

where $Y_{imp,i}$ is the imputed Y value for the i^{th} plot and ‘IMP’ refers to either tree-level RF or plot-level RF.

SAMPLE25, MA, WMA, and the tree-level and plot-level RF imputation methods were compared based on the overall means of the variables of interest in 2000 (see Equations 2 – 5). The five estimation methods were also compared based on their performance of estimating the four variables of interest by species groups ‘Douglas-fir’, ‘pine’, and ‘other.’

The basis of evaluation was accuracy, as expressed by the root mean square error (RMSE), and bias, calculated as the mean difference between the estimates (Equations 2 – 5) and the observed mean values (Equation 1) from $m = 200$ iterations of randomly splitting the data. Two hundred iterations were considered sufficient because other studies have found RMSE and bias to stabilize at around 200 iterations (e.g., Arner et al. 2004). Both RMSE and bias were expressed as percent of the observed mean for each variable of interest:

$$Bias \% = \frac{\sum_{i=1}^n \frac{(est_i - obs_i)}{m}}{\frac{\sum_{i=1}^n obs_i}{m}} * 100 \quad [6]$$

$$RMSE \% = \frac{\sqrt{\sum_{i=1}^n \frac{(est_i - obs_i)^2}{m}}}{\frac{\sum_{i=1}^n obs_i}{m}} * 100 \quad [7]$$

Results

The SAMPLE25 estimator provided virtually unbiased estimates for all four variables of interest. Its RMSE values ranged from 4.89% for SPH to 6.58% for BIOT. The MA estimates had a negative bias with values from -1.93% for VOL to -2.58% for SPH. The MA estimator provided very precise estimates with the bias contributing most to the RMSE. The WMA estimator reduced both bias and RMSE for BA and SPH. For WMA the bias of VOL and BIOT estimates was positive and the RMSE was larger than those for MA (Table 3).

Plot-level RF imputation resulted in small negative bias and smaller RMSE values than those of the MA estimator. In terms of RMSE, plot-level RF imputation only outperformed the WMA for VOL and BIOT (Table 3).

Table 3: Imputation results.

Method	BA		SPH		VOL		BIOT	
	% bias	% RMSE	% bias	% RMSE	% bias	% RMSE	% bias	% RMSE
SAMPLE25	0.05	5.29	-0.20	4.89	0.20	6.53	0.26	6.58
MA	-2.53	2.60	-2.58	2.63	-1.93	2.08	-1.97	2.12
WMA	0.59	1.03	-1.54	1.72	2.52	2.74	2.62	2.83
plot-level RF	-0.44	1.50	-0.73	2.52	-0.26	1.78	-0.22	1.66
tree-level RF	0.44	1.09	-0.60	1.31	0.43	1.36	0.42	1.35

Tree-level RF imputation produced a small positive bias in BA, VOL, and BIOT but a small negative bias in SPH. Its RMSE values were smaller than those of the MA and the plot-level RF imputation. Tree-level RF imputation outperformed the WMA estimates in terms of bias and RMSE for SPH, VOL, and BIOT. The variance contributed most to the RMSE for both tree- and plot-level imputation (Table 3).

By species group the SAMPLE25 estimator provided virtually unbiased results (0.62% or less). RMSE values ranged from 8.52% for ‘pine’ BA to 11.17% for ‘other’ BIOT (Table 4).

The MA estimator resulted in a larger negative bias for the four variables of interest for species group ‘pine’ which contributed most to the RMSE values of more than 9%. For the species group ‘Douglas-fir’ and ‘other,’ MA resulted in small bias with absolute values ranging from 0.30% to 1.17% and RMSE values ranging from 1.00% to 1.68% (Table 4).

WMA estimates were biased for all three species groups with the bias being largest for ‘pine.’ The bias contributed most to the RMSE values, which exceeded the RMSE values of the MA estimates and the RMSE values for ‘pine’ for the SAMPLE25 estimates (Table 4).

Table 4: Tree- and plot-level imputation results by species group.

Method	BA 'Douglas-fir'		BA 'pine'		BA 'other'		SPH 'Douglas-fir'		SPH 'pine'		SPH 'other'	
	% bias	% RMSE	% bias	% RMSE	% bias	% RMSE	% bias	% RMSE	% bias	% RMSE	% bias	% RMSE
SAMPLE25	0.41	9.94	0.29	8.52	-0.33	10.10	-0.39	10.21	0.49	10.04	-0.56	9.99
MA	0.40	1.14	-9.85	9.89	-0.70	1.18	-0.30	1.00	-9.28	9.35	0.61	1.15
WMA	6.63	6.79	-13.49	13.57	3.86	4.12	2.21	2.64	-14.51	14.61	4.95	5.14
plot-level RF	2.34	9.65	2.19	9.75	-4.01	8.25	2.05	11.01	2.01	10.36	-4.13	9.01
tree-level RF	-0.54	1.63	0.98	1.68	-0.39	1.21	-0.57	2.25	-1.54	2.80	-0.93	1.74

Method	VOL 'Douglas-fir'		VOL 'pine'		VOL 'other'		BIOT 'Douglas-fir'		BIOT 'pine'		BIOT 'other'	
	% bias	% RMSE	% bias	% RMSE	% bias	% RMSE	% bias	% RMSE	% bias	% RMSE	% bias	% RMSE
SAMPLE25	0.62	10.99	0.23	8.84	-0.15	11.05	0.56	10.68	0.23	8.62	0.02	11.17
MA	-0.77	1.50	-9.27	9.31	0.48	1.19	-1.17	1.68	-9.36	9.40	0.48	1.21
WMA	6.88	7.09	-11.89	11.98	5.59	5.82	6.06	6.28	-12.15	12.23	5.95	6.17
plot-level RF	2.00	9.94	1.94	10.36	-3.17	8.57	1.81	9.63	2.18	10.24	-3.04	8.39
tree-level RF	-0.84	2.05	3.74	4.02	-1.25	1.97	-3.37	3.76	8.36	8.54	-0.91	1.81

Plot-level RF imputations resulted in a smaller bias than with WMA for all species groups. However, RMSE values for RF exceeded those of WMA for all but ‘pine’ (Table 4).

Tree-level RF imputation outperformed SAMPLE25, WMA, and plot-level RF imputation in terms of RMSE. Compared to MA, tree-level RF imputation provided smaller RMSE values for ‘pine’ and slightly larger RMSE values for ‘Douglas-fir’ and ‘other’ (Table 4).

Discussion

The performance of the MA estimator in terms of the variance-bias trade-off was as expected. As in most other studies (e.g., Arner et al. 2004, Johnson et al. 2003, Van Deusen 2002), the large bias was found to be more than compensated for by the high precision. Hence, MA provided better estimates in terms of accuracy than SAMPLE25. MA is a temporal ‘midpoint’ estimator yielding biased estimates at the end of a time-series in the presence of trend (Roesch and Reams 1999). Giving more weight to the more recently measured panels resulted in the WMA estimator, which improved the estimates for BA and SPH in terms of bias and hence, also in terms of RMSE.

MA by species groups outperformed WMA in terms of both bias and RMSE. The larger weights applied for *P3* and *P4* for WMA increased the negative bias for species group ‘pine’ and resulted in large positive bias for all variables of interest for species groups ‘Douglas-fir’ and ‘other.’ This indicates that weights applied to the WMA which improve the MA for estimating BA, SPH, VOL, and BIOT do not necessarily improve the MA when the variables of interest are summarized by species group. Choosing appropriate weights for the WMA requires the knowledge of the trend inherent in the data. If the trend inherent in BA, SPH, VOL, and BIOT differs from the trend of the variables of interest summarized by species group, different weights need to be chosen for the WMA. Objective ways for choosing appropriate weights are still lacking. Panels that do not change much should receive larger weights than panels that exhibit a lot of change. Knowledge about change could possibly be acquired from remotely sensed data, growth models, or other information on, for example, fire or insect outbreaks.

Tree-level RF imputation outperformed MA in terms of bias and RMSE for estimating BA, SPH, VOL, and BIOT. This is due to the lag bias inherent in the MA estimator. Tree-level imputation attempts to update the tree data, which results in a smaller bias than that observed for MA. Compared to the WMA, which tries to adjust the lag bias of the MA estimator, the improvement of tree-level RF imputation is less pronounced. If the lag bias of the MA could be adjusted, MA might outperform RF tree-level imputation in terms of both bias and RMSE since the MA estimates are more precise than those of the tree-level RF imputation.

When the variables of interest were summarized by species groups, the MA slightly outperformed tree-level RF imputation in terms of RMSE for species groups ‘Douglas-fir’ and ‘other’ because the MA resulted in low bias for those variables. For species group ‘pine’ MA resulted in large bias and therefore tree-level RF imputation provided much better results for the variables of interest for species group ‘pine’ in terms of bias and RMSE.

Tree-level RF imputation outperformed plot-level imputation for estimating BA, SPH, VOL, and BIOT as well as for estimating the variables of interest summarized by species groups. The results of this study suggest that tree-level RF imputation should be preferred over plot-level RF imputation for estimating total BA, SPH, VOL, and BIOT or for estimating BA, SPH, VOL, and BIOT by species group. The same considerations for choosing single-tree growth models over whole-stand growth models probably apply for choosing tree-level NN imputation over plot-level NN imputation and depend mainly on the demands of the user.

In this study, tree-level variables were imputed using reference trees irrespective of whether the tree species of reference and target trees matched. Imputing only within tree species or species group might improve the results for tree species such as Douglas-fir, ponderosa pine, grand fir, lodgepole pine, white fir, and western hemlock which occur frequently in the data set. However, results for rare tree species would definitely degrade with decreasing number of possible reference trees. Overall results could possibly be improved by imputing tree-variables for frequent tree species within tree species but using the complete reference data set for rare tree species.

Conclusions

This study has shown that tree-level RF imputation has the potential to provide better results in terms of bias and accuracy for estimating plot-level attributes such as BA, SPH, VOL, and BIOT than can be achieved with the SAMPLE25, MA, and WMA estimators, or plot-level RF imputation.

Giving more weight to most recently measured panels by using a WMA improved the estimates for BA, SPH, VOL, and BIOT compared to the MA estimates. When the variables of interest were summarized by species group, MA outperformed WMA in terms of bias and accuracy. More research is warranted for finding objective methods for choosing appropriate weights.

Tree-level RF imputation outperformed MA and WMA in terms of bias and accuracy when BA, SPH, and VOL were estimated. When the variables of interest were summarized by species group, MA provided slightly better estimates in terms of accuracy than tree-level RF imputation.

Tree-level imputation outperformed plot-level imputation. This might be due to the fact that tree-level NN imputation requires more information and is based on a more detailed representation of the stand than plot-level imputation.

The results of the tree-level NN imputation methods tested in this study provide a good argument to further develop the application of tree-level NN imputation techniques for estimating current forest attributes from paneled inventory data.

Acknowledgements

We thank Janet Ohmann and Matt Gregory for sharing their ancillary data from the NLCD and DAYMET with us and for their insights on the data. We appreciate the help that Jim Alegria, Carol Apple, Bob Brown, and Melinda Moeur provided in obtaining national forest data, and thank Kurt Campbell for assistance with volume and biomass equations. We also thank Frank Roesch, Steen Magnussen, and Emilie Grossmann for their helpful review comments.

Literature Cited

- Arner, S.L., J.A. Westfall, and C.T. Scott. 2004. Comparison of annual inventory designs using forest inventory and analysis data. *Forest Science* 50(2): 88-203.
- Barrett, T.M. 2006. Optimizing efficiency of height modeling for extensive forest inventories. *Canadian Journal of Forest Research* 36:2259-2269.
- Breiman, L. 2001. Random forests. *Machine Learning* 45(1): 5-32.
- Burkhardt, H.E. 1992. Tree and stand models in forest inventory. P. 164-170 in *Proceedings of Ilvessalo Symposium on National Forest Inventories*. IUFRO S4.02. Finnish Forest Research Institute Research Papers 444, Helsinki, Finland.
- Crookston, N.L. and A.O. Finley. 2008. yaImpute: An R package for kNN Imputation. *Journal of Statistical Software* 23(10):1-16.
- Curtis, R.O. and D.M. Hyink. 1985. Data for growth and yield models. P. 1-5 in *Proceedings – Growth and yield and other mensurational tricks: A regional technical conference*, Van Hooser, D.D. and N. Van Pelt (eds.). USDA Forest Service General Technical Report INT-GTR-193, Ogden, UT.
- Eskelson, B.N.I., H. Temesgen, and T.M. Barrett. 2009. Estimating current forest attributes from paneled inventory data using plot-level imputation: a study from the Pacific Northwest. *Forest Science* 55(1):64-71.
- Environmental Systems Research Institute [ESRI]. 1991. Cell-based modeling with GRID. ESRI Inc., US, Redlands, California.
- Fehrmann, L., A. Lehtonen, C. Kleinn, and E. Tomppo. 2008. Comparison of linear and mixed-effect regression models and a k-nearest neighbour approach for estimation of single-tree biomass. *Canadian Journal of Forest Research* 38:1-9.
- Gartner, D. and G.A. Reams. 2001. A comparison of several techniques for estimating the average volume per acre for multipanel data with missing panels. P. 76-81 in *Proceedings of the second annual Forest Inventory and Analysis Symposium*, Reams, G.A., R.E. McRoberts, and P.C. Van Deusen (eds.). USDA Forest Service General Technical Report SRS-GTR-47, Asheville, NC.
- Gartner, D., and G.A. Reams 2002. A comparison of several techniques for imputing tree level data. P. 103-108 in *Proceedings of the Third Annual Forest Inventory and*

- Analysis Symposium*, McRoberts, R.E., G.A. Reams, P.C. Van Deusen, and J.W. Moser (eds.). USDA Forest Service General Technical Report NC-GTR-230, St. Paul, MN.
- Homer, C., C. Huang, L. Yang, B. Wylie, and M. Coan. 2004. Development of a 2001 National Landcover Database for the United States. *Photogrammetric Engineering and Remote Sensing* 70(7):829-840.
- Hudak, A.T., N.L. Crookston, J.S. Evans, D.E. Hall, and M.J. Falkowski. 2008. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. *Remote Sensing of Environment* 112(5): 2232-2245.
- Johnson, D.S., M.S. Williams, and R.L. Czaplewski. 2003. Comparison of estimators for rolling samples using forest inventory and analysis data. *Forest Science* 49(1):50-63.
- Kauth, R.J., and G.S. Thomas. 1976. The tasseled cap—a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. PP. 4B41-4B51 in *Proceedings of the symposium on machine processing of remotely sensed data*, Purdue University. West Lafayette, IN.
- Korhonen, K.K., and A. Kangas. 1997. Application of nearest-neighbour regression for generalizing sample tree information. *Scandinavian Journal of Forest Research* 12:97-101.
- Lessard, V.C., R.E. McRoberts, and M.R. Holdaway. 2001. Diameter growth models using Minnesota forest inventory and analysis data. *Forest Science* 47(3):301-310.
- Max, T.A., H.T. Schreuder, J.W. Hazard, J. Teply, and J. Alegria. 1996. The Region 6 vegetation inventory and monitoring system. USDA Forest Service General Technical Report PNW-RP-493. 22 p.
- McRoberts, R.E. 2001. Imputation on model-based updating techniques for annual forest inventories. *Forest Science* 47(3):322-330.
- Ohmann, J.L., and M.J. Gregory. 2002. Predictive mapping of forest composition and structure with direct gradient analysis and nearest-neighbor imputation in coastal Oregon, U.S.A. *Canadian Journal of Forest Research* 32:725-741.
- Patterson, P.L., and G.A. Reams. 2005. Combining panels for forest inventory and analysis estimation. P. 69-74 in *The enhanced forest inventory and analysis program—National sampling and estimation procedures*, Bechtold, W.A., and P.L. Patterson (eds.). USDA Forest Service General Technical Report SRS-GTR-80. Asheville, NC.
- Roesch, F.A., and G.A. Reams. 1999. Analytical alternatives for an annual inventory system. *Journal of Forestry* 97(12):33-37.
- Sironen, S., Kangas, A., Maltamo, M., and J. Kangas. 2001. Estimating individual tree growth with k-nearest neighbour and k-most similar neighbour methods. *Silva Fennica* 35(4):453-467.
- Sironen, S., A. Kangas, M. Maltamo, and J. Kangas. 2003. Estimating individual tree growth with nonparametric methods. *Canadian Journal Forest Research* 33:444-449.
- Sironen, S., A. Kangas, M. Maltamo, and J. Kangas. 2008. Localization of growth estimates using non-parametric imputation methods. *Forest Ecology and Management* 256:674-684.
- Thornton, P.E. and S.W. Running. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. *Agricultural and Forest Meteorology* 93:211-228.
- Thornton, P.E., S.W. Running, and M.A. White. 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology* 190:214-251.

- USDA 2000. Western Oregon 1995-1997, August 15th, 2000. A CD available online from USDA/ FS PNW Research at <http://www.fs.fed.us/pnw/fia/>; last accessed Sep.10, 2008.
- Van Deusen, P.C. 1996. Incorporating predictions into an annual forest inventory. *Canadian Journal of Forest Research* 26(9):1709-1713.
- Van Deusen, P.C. 2002. Comparison of some annual forest inventory estimators. *Canadian Journal of Forest Research* 32:1992-1995.