Imputing Mean Annual Change to Estimate Current Forest Attributes

Bianca N. I. Eskelson, Tara M. Barrett and Hailemariam Temesgen

When a temporal trend in forest conditions is present, standard estimates from paneled forest inventories can be biased. Thus methods that use more recent remote sensing data to improve estimates are desired. Paneled inventory data from national forests in Oregon and Washington, U.S.A., were used to explore three nearest neighbor imputation methods to estimate mean annual change of four forest attributes (basal area/ha, stems/ha, volume/ha, biomass/ha). The randomForest imputation method outperformed the other imputation approaches in terms of root mean square error. The imputed mean annual change was used to project all panels to a common point in time by multiplying the mean annual change with the length of the growth period between measurements and adding the change estimate to the previously observed measurements of the four forest attributes. The resulting estimates of the mean of the forest attributes at the current point in time outperformed the estimates obtained from the national standard estimator.

1   Introduction

Keeping national inventories of forests updated to reflect current conditions poses substantial logistical and accuracy issues (Gillis and Leckie 1996). In paneled inventory systems the plot network is systematically divided into panels with rotating panels being measured on an annual basis (Reams et al. 2005); in a 10 year panel inventory, for example, one-tenth of all plots are measured each year and plots of each panel are remeasured every 10 years. In theory, a paneled inventory system can provide current information when only the most recent panel is used (Reams and Van Deusen 1999), but in practice, most forestry applications are at a spatial scale that require combining field plots from multiple years to achieve sufficient information (McRoberts 2001, Tomppo et al. 2008). However, combining field plots from multiple years to estimate current conditions can cause a lag bias when forest conditions are changing over time.

In the United States (US), the national inventory of forests is collected with a panel system by the Forest Inventory and Analysis (FIA) program of the US Forest Service. The FIA default estimator is a moving average (MA) approach (Bechtold and Patterson 2005) which is known to result in biased estimates when trend is present (Van Deusen 2002). In the western US, the problem of lag bias is exacerbated by a relatively long remeasurement interval (10 years), shifts in forest management in response to altered economic and social conditions, changing climate, and a high but variable disturbance rate from wildfire, disease, and insects. Thus there is interest in using remote sensing information to reduce lag bias in estimates of current forest conditions. Combining remote sensing and other ancillary data with field plots has become common for improving forest inventory information (Tomppo et al. 2008).

Combining field and remote sensing data, nearest neighbor (NN) imputation is often used for mapping (Finley and McRoberts 2008, Katila and Tomppo 2002, Ohmann and Gregory 2002), and these methods typically impute point-in-time plot attributes. In contrast, when using imputation to update paneled inventory data, it is possible to impute mean annual change (MAC) for a plot. For example, Arner et al. (2004) estimated mean

annual net volume change using MA approaches and sampling with partial replacement approaches and McRoberts (2001) imputed the difference in basal area between two measurements to plots with missing measurements to update basal area for plots measured in previous years to the current point in time. There are few studies in the western US that examine alternatives to the MA (e.g., Eskelson et al. 2009) and none that we are aware of that impute MAC rather than point-in-time measurements.

The objectives of this study are to: 1) use paneled data from the Pacific Northwest (PNW) to estimate mean annual change (MAC) of forest attributes using three nearest neighbor imputation methods; and 2) to estimate current forest attributes from paneled inventory data by updating the most recent measurement with imputed MAC. The results are compared with the estimates obtained from the MA estimator and the data from the current panel.

2   Methods

2.1 Data

In this study, 618 primary sampling units (PSU) from six national forests that were collected as part of the Pacific Northwest Region's Current Vegetation Survey (CVS) of the US Forest Service were used. The particular national forests, sampled between 1993 and 1997 (occasion 1 measurement) and remeasured in 2000, were the Colville (28), Mt. Hood (111), Ochoco (82), Rogue River (70), Wallowa-Whitman (199), and Winema (128) (Fig. 1).

Panel data is a special case of inventory data with measurements taken at different times. A panel system was imitated with the available data by randomly assigning 25% of the PSUs to panel 4 (*P4*; 154 PSUs) and the remaining 75% of the PSUs (464 PSUs) to panel 1 (*P1*), panel 2 (*P2*), and panel 3 (*P3*) based on their year of occasion 1 measurement (Table 1).

In the CVS inventory, circular PSUs (1 ha in size) are established on a regular grid with square spacing (5.47 km). Each circular PSU is subsampled by a satellite system of five second-

Fig. 1. Geographical distribution of the six national forests in Washington and Oregon.

Table 1. Number of primary sampling units (PSU) measured by year of the occasion 1 measurement and corresponding panel assignment. All PSUs (n=618) were remeasured in 2000.

Occasion 1 measurement	# of PSU	Assigned panel
1993	7	1
1994	229	1
1995	223	2
1996	158	3
1997	1	3

ary sampling units (SSU). Each SSU consists of three fixed-area circular, nested plots with radii of 3.6 m, 7.3 m, and 15.6 m in which trees with diameter at breast height (DBH in cm) smaller

than 12.7 cm, with 12.7 cm ≤ DBH > 33 cm, and DBH greater or equal to 33 cm are measured, respectively. For a detailed description of the CVS inventory see Max et al. (1996). Live trees with DBH of 12.7 cm or larger were used in this study and height models developed in Barrett (2006) were employed to fill in missing heights (HT in m). Gross cubic-meter volume and total gross oven dry weight biomass were calculated with volume and biomass equations from the US Forest Service (USDA 2000). For each PSU, basal area in m^2 per ha (BA), stems per ha (SPH), volume in m^3 per ha (VOL), and biomass in tons per ha (BIOT) were calculated and summarized (Table 2). MAC for BA, SPH, VOL, and BIOT were calculated by dividing the difference of the observed values in 2000 and the observed value at the occasion 1 measurement by the growth period

Table 2. Summary of variables at the primary sampling units (n = 618) in 2000.

Variable	Minimum	Mean	Maximum	Standard deviation
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
ln(Annual precipitation) (mm) (scaled * 100)	577	683	817	48
Annual temperature (°C) (scaled * 100)	60	579	1067	166

length (GPL) between the two measurements.

Thematic Mapper (TM) images from 2000 from the national land cover database 2001 (USDI / USDA 2009) were used as ancillary data. This data set includes normalized imagery bands 1 to 5 and band 7 (TM1, TM2, TM3, TM4, TM5, TM7), three commonly used band ratios (band 4 to band 3, band 5 to band 4, and band 5 to band 7), the normalized difference vegetation index (NDVI), the Tasseled Cap (TC) transformations of the 6 axes (Kauth and Thomas 1976), and a model of percent canopy cover. Normalization methods and development of derived layers are described in Homer et al. (2004). Values for each plot were calculated by overlaying the TM attributes, which had a 30 m resolution, with the center points of the five SSUs and calculating a mean of the five values.

In addition to the satellite data, the following climate and topography variables were used as ancillary data (Table 2): Natural logarithm of annual precipitation (mm) and mean annual temperature (°C) [Data source: DAYMET Daily Surface Weather Data and Climatological Summaries (Thornton et al. 1997, Thornton and Running 1999)], Elevation (EL in m) and its transformations EL^2 and ln(EL) [Data source: CVS inventory], slope (%), and the transformations 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)].

2.2 Nearest Neighbor Imputation

Nearest neighbor (NN) imputation methods are donor-based methods. Variables of interest are those forest attributes that are only measured on a subset of PSUs (e.g., MAC of BA, SPH, VOL, and BIOT). Ancillary variables are the attributes that are measured on all PSUs. In this study, satellite, climate, and topography data as well as the most recent measurements of BA, SPH, VOL, and BIOT, which were taken at measurement occasion 1 (between 1993 and 1997), constitute the available ancillary data. Reference data are the PSUs for which both variables of interest and ancillary variables are available (PSUs in *P4*). Target data are the PSUs for which only the ancillary variables are available (PSUs in *P1*, *P2*, and *P3*). The reference PSUs constitute the pool of potential PSUs which could be selected to impute the MAC data for the target PSUs.

The most similar neighbor (MSN) method (Moeur and Stage 1995), the gradient nearest neighbor (GNN) method (Ohmann and Gregory 2002), and the randomForest (RF) method (Crookston and Finley 2008) have been shown to provide reasonable imputation results for forest attributes (Moeur and Stage 1995, Temesgen et al. 2003, LeMay and Temesgen 2005, Hudak et al. 2008) and for mapping forest composition and structure (Ohmann and Gregory 2002, Ohmann et al. 2007). MSN, GNN, and RF were conducted using the yaImpute R package version 1.0–9 (Crookston and Finley 2008). The

similarity between reference and target PSUs is defined using a weighted Euclidean distance for MSN and GNN:

$$D_{ij}^2 = (X_i - X_j)W(X_i - X_j)' \tag{1}$$

where W is the weight matrix, X_i is a vector of standardized values of the ancillary variables for the ith target PSU; and X_j is a vector of standardized values of ancillary variables for the jth reference PSU. The ancillary variables for both target and reference PSUs were standardized using the mean and variance of the ancillary variables of the reference PSUs resulting in standardized variables with zero mean and unit variance. For MSN, the weight used is $W = \Gamma \Lambda^2 \Gamma'$, where Γ is the matrix of standardized canonical coefficients for the ancillary variables and Λ^2 is the diagonal matrix of squared canonical correlations between ancillary attributes and variables of interest (Moeur and Stage 1995).

For the gradient nearest neighbor method (GNN) the weights are assigned by employing a projected ordination of the ancillary data based on canonical correspondence analysis (CCA) (Ohmann and Gregory 2002).

The RF method is an extension of classification and regression tree (CART) methods (Breiman 2001). The data and variables are randomly and iteratively sampled to generate a forest of classification and regression trees. If two PSUs tend to end up in the same terminal nodes in a forest of classification and regression trees, they are considered to be similar. The RF distance measure is one minus the proportion of trees where a target PSU is in the same terminal node as a reference PSU (Crookston and Finley 2008, Hudak et al. 2008).

2.3 Estimation Procedures

BA, SPH, VOL, and BIOT were the variables of interest (Y). The observed mean value of Y in 2000 ($\bar{Y}_{OBS}$), based on the observed Y values from all 618 PSUs, was used as the best available estimate of the true mean of Y:

$$\bar{Y}_{OBS} = \sum_{i=1}^{n} Y_i / n \tag{2}$$

where Y_i is the observed Y value of the ith PSU in 2000 and $n = 618$.

The 618 PSUs were randomly split into $P4$ (154 PSUs) and $P1$, $P2$, and $P3$ (remaining 464 PSUs) 500 times. For each of the 500 realizations of randomly splitting the data, the following estimators were used to estimate the mean value of the variables of interest for the year 2000.

Using the data from $P4$ only, the mean value of Y for the year 2000 (SAMPLE25 estimator) was estimated as:

$$\hat{\bar{Y}}_{SAMPLE25} = \sum_{i:Y_i \in P4} Y_i / n_4 \tag{3}$$

where Y_i is the observed Y value of the ith PSU, and n_4 is the number of PSUs in $P4$.

The MA estimator, the FIA default method, was also used to estimate the mean value of Y for the year 2000:

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

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 (e.g., $\bar{Y}_{t-3,i} = \sum_{i:Y_i \in P1} Y_i / n_1$, where n_1 is the number of PSUs in $P1$), with $t-3$, $t-2$, and $t-1$ being the years 1993/1994, 1995, and 1996/1997, respectively. In the following, this MA(4) estimator will be referred to as MA.

Instead of filling in the missing values for $P1$, $P2$, and $P3$ with their previous measurements, as was done in the MA calculation, MSN, GNN, and RF were explored to impute the MAC of the variables of interest for the PSUs in $P1$, $P2$, and $P3$. The R package yaImpute (Crookston and Finley 2008) was used to impute MAC. The PSUs in $P4$ were the reference data in the imputation process. The imputed MAC was then used to update the variables of interest for each PSU in $P1$, $P2$, and $P3$ to the year 2000 by multiplying it with the GPL and adding it to the occasion 1 measurement as follows:

$$Y_{imp,i} = Y_{t-x,i} + MAC_{imp,i} * GPL_i \tag{5}$$

where $Y_{imp,i}$ is the imputed Y value for the ith PSU in 2000 and $Y_{t-x,i}$ is the observed Y value for the

*i*th PSU at time *t–3* (1993/1994), *t–2* (1995), and *t–1* (1996/1997) for *P1*, *P2*, and *P3*, respectively. $MAC_{imp,i}$ is the imputed MAC for the *i*th PSU with *imp* referring to the NN imputation method used. GPL_i is the growth length period between

the occasion 1 measurement and the remeasurement in 2000.

The mean value of *Y* for the year 2000 was then estimated as follows:

$$\hat{\bar{Y}}_{IMP} = \left[\sum_{i:Y_i \in P1} Y_{imp,i} + \sum_{i:Y_i \in P2} Y_{imp,i} + \sum_{i:Y_i \in P3} Y_{imp,i} + \sum_{i:Y_i \in P4} Y_{t,i} \right] / n \qquad (6)$$

where IMP refers to the NN imputation method used and $Y_{imp,i}$ is the imputed *Y* value for the *i*th PSU as described in Eq. 5.

Two sets of ancillary variables were tested for the NN imputation methods: The first set included the available climate, topography, and satellite data (Data set A) and the second set consisted of the previous measurements of the variables of interest that were taken at measurement occasion 1 from 1993 to 1997 (Data set B: BAocc1, SPHocc1, VOLocc1, BIOTocc1).

SAMPLE25, MA, and the three imputation methods were compared based on the estimated overall means of the variables of interest in 2000 (see Eqs. 3, 4, and 6). The basis of evaluation was accuracy, as expressed by the root mean square error (RMSE), and bias calculated as the mean difference between the estimated and the observed (Eq. 2) mean values. RMSE and bias were approximated using a random sample of $m = 500$ realizations of splitting the data. Both RMSE and bias were expressed as percent of the observed mean for each variable of interest:

$$\text{Bias \%} = \frac{\sum_{j=1}^{m} \dfrac{\left(\hat{\bar{Y}}_j - \bar{Y}_{OBS} \right)}{m}}{\bar{Y}_{OBS}} * 100 \qquad (7)$$

$$\text{RMSE \%} = \frac{\sqrt{\sum_{j=1}^{m} \dfrac{\left(\hat{\bar{Y}}_j - \bar{Y}_{OBS} \right)^2}{m}}}{\bar{Y}_{OBS}} * 100 \qquad (8)$$

where $\hat{\bar{Y}}_j$ is the estimate of $\bar{Y}$ for the *j*th iteration of data splitting using either SAMPLE25, MA, or the NN imputation methods. Two-sided t-tests were performed to examine whether the esti-

mated bias (in original units) of the SAMPLE25, MA, and the NN imputation methods was significantly different from zero, and the p-values were reported.

The MAC estimates based on MSN, GNN, and RF imputation were compared to the observed MAC also using the approximated RMSE and bias from $m = 500$ iterations of randomly splitting the data. Two-sided t-tests were performed to examine whether the estimated bias (in original units) of the three NN imputation methods using data sets A and B for imputing MAC was significantly different from zero, and the p-values were reported.

3 Results

MSN imputation provided similar results for MAC estimates for both sets of ancillary variables with the variance contributing most to the RMSE. When BAocc1, SPHocc1, VOLocc1, and BIOTocc1 were used as ancillary variables, the p-values of the two-sided t-test testing the unbiasedness of the estimator ranged from 0.7041 to 0.9453 and the p-values for the estimates based on climate, topography, and satellite data ranged from 0.1556 to 0.4581 indicating that MSN imputation provided unbiased MAC estimates with both ancillary data sets (Table 3). GNN estimates of MAC were significantly biased ($p < 0.0001$) and the large bias ($>66\%$) contributed most to the RMSE. Using climate, topography, and satellite data as ancillary variables resulted in smaller bias and hence, smaller RMSE values than using BAocc1, SPHocc1, VOLocc1, and BIOTocc1 as ancillary variables (Table 3). In terms of RMSE, RF using climate, topography, and satellite data as ancillary variables provided the best estimates of

Table 3. Bias with p-value in parentheses and RMSE of mean annual change (MAC) of BA (basal area/ha), SPH (stems/ha), VOL (volume/ha), and BIOT (biomass/ha). Data set A comprises climate, topography, and satellite data. Data set B comprises occasion 1 measurements of the variables of interest.

Method	Data	MAC in BA		MAC in SPH		MAC in VOL		MAC in BIOT	
		Bias%	RMSE%	Bias%	RMSE%	Bias%	RMSE%	Bias%	RMSE%
MSN	A	1.40 (0.1556)	22.09	−1.89 (0.3180)	42.26	0.82 (0.4581)	24.81	0.93 (0.3717)	23.34
GNN	A	68.60 (<0.0001)	79.72	155.10 (<0.0001)	171.37	72.70 (<0.0001)	84.90	66.89 (<0.0001)	79.47
RF	A	−2.86 (0.0007)	16.13	−2.82 (0.0770)	35.67	−5.67 (<0.0001)	19.42	−5.80 (<0.0001)	17.85
MSN	B	−0.16 (0.8364)	22.02	−1.16 (0.9453)	42.69	0.52 (0.7041)	24.86	0.30 (0.7485)	23.59
GNN	B	98.34 (<0.0001)	136.74	297.26 (<0.0001)	475.71	88.54 (<0.0001)	121.62	95.48 (<0.0001)	131.76
RF	B	−0.33 (0.7205)	20.67	−2.5 (0.1480)	38.58	0.28 (0.7930)	23.92	−0.65 (0.5168)	22.36

MAC. However, these were biased for BA, VOL, BIOT (p<0.0007) and suggestive, but inconclusively unbiased for SPH (p=0.0770). RF using BAocc1, SPHocc1, VOLocc1, and BIOTocc1 as ancillary data provided unbiased MAC estimates (p>0.1) for all four variables of interest and smaller RMSE values than those achieved by either MSN or GNN imputation (Table 3).

For estimating current BA and SPH the RMSE values of MA were about half the size of those observed for SAMPLE25. For estimating current VOL and BIOT the RMSE values for MA were about a third of those observed for SAMPLE25. SAMPLE25 results were unbiased (p>0.5). MA results were precise but biased (p<0.0001) and the bias contributed most to the RMSE. The opposite was true for the unbiased SAMPLE25 estimates, where the variance contributed most to the RMSE (Table 4).

MSN resulted in biased estimates of current BA, SPH, VOL, and BIOT (p<0.0047) when BAocc1, SPHocc1, VOLocc1, and BIOTocc1 were used as ancillary variables. When climate, topography, and satellite data were used as ancillary variables, estimates were biased for BA (p=0.0021), unbiased for SPH and VOL (p>0.1), and suggestive, but inconclusively unbiased for BIOT (p=0.0460). For both sets of ancillary variables, the MSN estimates were similar in terms

of RMSE and outperformed the MA estimates in terms of both bias and RMSE (Table 4).

For the GNN estimates of current forest attributes, bias contributed most to RMSE. When BAocc1, SPHocc1, VOLocc1, and BIOTocc1 were used as ancillary variables, the bias and hence the RMSE was larger than for the ancillary variable set including climate, topography, and satellite data. For both sets of ancillary variables, bias and RMSE were larger than for any of the other estimators (Table 4).

RF estimates of the current variables of interest were biased (p<0.0016) when BAocc1, SPHocc1, VOLocc1, and BIOTocc1 were used as ancillary variables. When climate, topography, and satellite data were used as ancillary variables, RF provided suggestive, but inconclusively unbiased (p=0.0421) and unbiased (p=0.9975) estimates of current BA and SPH, respectively, and biased estimates of VOL and BIOT (p<0.0001). RMSE values were smallest when climate, topography, and satellite data were used as ancillary variables. For both sets of ancillary variables, RF imputation outperformed the MA estimates both in terms of bias and RMSE (Table 4).

RF using climate, topography, and satellite data as ancillary variables provided the best results overall in terms of bias and RMSE for estimating current BA, SPH, VOL, and BIOT, followed by

Table 4. Bias with p-value in parentheses and RMSE of mean BA (basal area/ha), SPH (stems/ha), VOL (volume/ha), and BIOT (biomass/ha) in year 2000. Data set A comprises climate, topography, and satellite data. Data set B comprises occasion 1 measurements of the variables of interest.

Method	Data	BA		SPH		VOL		BIOT	
		Bias%	RMSE%	Bias%	RMSE%	Bias%	RMSE%	Bias%	RMSE%
SAMPLE25		−0.02 (0.9344)	5.29	0.13 (0.5769)	5.05	−0.08 (0.7909)	6.59	−0.06 (0.8423)	6.67
MA		−2.54 (<0.0001)	2.60	−2.63 (<0.0001)	2.68	−1.92 (<0.0001)	2.06	−1.98 (<0.0001)	2.12
MSN	A	0.22 (0.0021)	1.59	0.07 (0.4254)	2.03	0.11 (0.1809)	1.89	0.17 (0.0460)	1.86
GNN	A	5.40 (<0.0001)	6.19	7.83 (<0.0001)	8.63	6.01 (<0.0001)	6.97	5.82 (<0.0001)	6.84
RF	A	−0.10 (0.0421)	1.12	0.00 (0.9975)	1.72	−0.41 (<0.0001)	1.42	−0.40 (<0.0001)	1.35
MSN	B	0.26 (0.0002)	1.54	0.31 (0.0008)	2.07	0.23 (0.0047)	1.83	0.26 (0.0012)	1.81
GNN	B	7.66 (<0.0001)	10.38	15.02 (<0.0001)	23.77	7.16 (<0.0001)	9.76	8.16 (<0.0001)	11.10
RF	B	0.29 (<0.0001)	1.47	0.27 (0.0015)	1.88	0.29 (0.0003)	1.77	0.25 (0.0012)	1.72
PSU-level RF	B	−0.30 (<0.0001)	1.58	−0.85 (<0.0001)	2.78	−0.06 (0.4381)	1.90	−0.09 (0.2476)	1.79

MSN and RF using BAocc1, SPHocc1, VOLocc1, and BIOTocc1 as ancillary variables in terms of RMSE (Table 4).

In a previous study, RF imputation using BAocc1, SPHocc1, VOLocc1, and BIOTocc1 as ancillary data for directly imputing current BA, SPH, VOL, and BIOT provided the most accurate estimates compared to the SAMPLE25 and MA estimators and MSN and GNN imputation. The results of this PSU-level RF imputation are provided in Table 4. See Eskelson et al. (2009) for details. The RF imputation in this study outperforms the PSU-level RF imputation from the previous study in terms of RMSE for BA, SPH, VOL, and BIOT and in terms of bias for BA and SPH for both sets of ancillary data. For both sets of ancillary data MSN imputation in this study outperforms the PSU-level RF imputation from the previous study in terms of both RMSE and bias for SPH (Table 4).

4 Discussion

After the start of the second inventory cycle, MAC can be estimated using remeasured plots (Arner et al. 2004). Considering the year 2000 as the start of the second inventory cycle, MAC was estimated in this study. The results of the three NN imputation methods that were explored to impute MAC of forest attributes showed the same pattern that was observed in an earlier study where the same three NN imputation methods were used to impute BA, SPH, VOL, and BIOT (see Eskelson et al. 2009). RF imputation provided the best results in terms of RMSE followed by MSN. The results of this study suggest that GNN PSU-level imputation is not adequate to impute MAC of forest attributes. This might be due to the fact that the CCA in the GNN procedure requires the use of environmental factors for the ordination (Ohmann and Gregory 2002) which might not be picked up in the ancillary data that was used for the imputation in this study.

The performance of using MAC of forest attributes of one inventory cycle to predict MAC for the next inventory cycle could not be tested in this study since only one remeasurement of the PSUs was available for the CVS data. This is an area of research that should be pursued as soon as multiple remeasurements of the FIA annual inventory are available in the western US.

When BA, SPH, VOL, and BIOT were updated to the year 2000 using MAC estimates obtained from MSN and RF imputation, the estimates of the mean BA, SPH, VOL, and BIOT in 2000 outperformed those of the SAMPLE25 and MA estimators in terms of accuracy. These results indicate that updating the variables of interest for unmeasured PSUs to the current point in time using estimated MAC from MSN or RF imputation should be preferred over using the SAMPLE25 or MA estimators for estimating the current forest attributes.

The estimates of mean BA, SPH, VOL, and BIOT in 2000 using MAC estimates obtained from RF imputation also outperformed the estimates from PSU-level RF imputation that was used to directly impute BA, SPH, VOL, and BIOT in 2000 (see Eskelson et al. 2009). This is due to the fact that the approach employing the imputed MAC estimates makes use of the previously observed measurements directly, whereas the PSU-level RF imputation only makes use of the previous measurements indirectly by using them as ancillary data. Adding a multiple of estimated MAC to the previously observed measurement will result in current estimates that will be close to the actual values even if the estimated MAC values were not perfect. If current BA, SPH, VOL, and BIOT are imputed directly as was done in Eskelson et al. (2009), the imputed values can be either close to the actual values or they can be completely different. The results of this study suggest that the approach using imputed MAC values should be preferred over directly imputing current BA, SPH, VOL, and BIOT.

Acknowledgments

References

Arner, S.L., Westfall, J.A. & Scott, C.T. 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.

Bechtold, W.A. & Patterson, P.L. (eds.). 2005. The enhanced forest inventory and analysis program – national sampling design and estimation procedures. USDA Forest Service General Technical Report SRS-80. 85 p.

Breiman, L. 2001. Random forests. Machine Learning 45(1): 5–32.

Crookston, N.L. & Finley, A.O. 2008. yaImpute: an R package for kNN imputation. Journal of Statistical Software 23(10): 1–16.

Environmental Systems Research Institute [ESRI]. 1991. Cell-based modeling with GRID. ESRI Inc., US, Redlands, California.

Eskelson, B.N.I., Temesgen, H. & Barrett, T.M. 2009. Estimating current forest attributes from paneled inventory data using plot-level imputation: a study from the Pacific Northwest. Forest Science 5(1):64–71.

Finley, A.O. & McRoberts, R.E. 2008. Efficient k-nearest neighbor searches for multi-source forest attribute mapping. Remote Sensing of Environment 112(5): 2203–2211.

Gillis, M.D. & Leckie, D.G. 1996. Forest inventory update in Canada. The Forestry Chronicle 72(2): 138–156.

Homer, C., Huang, C., Yang, L., Wylie, B. & Coan, M. 2004. Development of a 2001 national landcover database for the United States. Photogrammetric Engineering and Remote Sensing 70(7): 829–840.

Hudak, A.T., Crookston, N.L., Evans, J.S., Hall, D.E., & Falkowski, M.J. 2008. Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. Remote Sensing of Environment 112(5): 2232–2245. Corrigendum in Remote Sensing of Environment 113(1): 289–290.

Katila, M. & Tomppo, E. 2002. Stratification by ancillary data in multisource forest inventories employing k-nearest neighbor estimation. Canadian Journal of Forest Research 32(9): 1548–1561.

Kauth, R.J. & Thomas, G.S. 1976. The tasseled cap – a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. In: Proceedings of the symposium on machine processing of remotely sensed data, Purdue University. West Lafayette, IN. p. 4B41–4B51.

LeMay, V. & Temesgen, H. 2005. Comparison of nearest neighbor methods for estimating basal area and stems per hectare using aerial auxiliary variables. Forest Science 51(2): 109–199.

Max, T.A., Schreuder, H.T., Hazard, J.W., Teply, J. & Alegria, J. 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.

Moeur, M. & Stage, A.R. 1995. Most similar neighbor: an improved sampling inference procedure for natural resource planning. Forest Science 41: 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, U.S.A. Canadian Journal of Forest Research 32: 725–741.

— , Gregory, M.J. & Spies, T.A. 2007. Influence of environment, disturbance, and ownership on forest vegetation of coastal Oregon. Ecological Applications 17(1): 18–33.

Reams, G.A. & Van Deusen, P.C. 1999. The southern annual forest inventory system. Journal of Agricultural, Biological, and Environmental Statistics 4(4): 345–359.

— , Smith, W.D., Hansen, M.H., Bechtold, W.A., Roesch, F.A. & Moisen, G.G. 2005. The forest inventory and analysis sampling frame. In: Bechtold, W.A. & Patterson, P.L. (eds.). The enhanced forest inventory and analysis program – national sampling design and estimation procedures. USDA Forest Service General Technical Report SRS-80. p. 11–16.

Temesgen, H., LeMay, V.M., Froese, K.L. & Marshall, P.L. 2003. Imputing tree-lists from aerial attributes for complex stands of south-eastern British Columbia. Forest Ecology and Management 177: 277–285.

Thornton, P.E. & Running, S.W. 1999. An improved algorithm for estimating incident daily solar radiation from measurements of temperature, humidity, and precipitation. Agricultural and Forest Meteorology 93: 211–228.

— , Running, S.W. & White, M.A. 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. Journal of Hydrology 190: 214–251.

Tomppo, E., Olsson, H., Ståhl, G., Nilsson, M., Hagner, O. & Katila, M. 2008. Combining national forest inventory field plots and remote sensing data for forest databases. Remote Sensing of Environment 112(5): 1982–1999.

Van Deusen, P.C. 2002. Comparison of some annual forest inventory estimators. Canadian Journal of Forest Research 32: 1992–1995.

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/. [Cited 10 Sep 2008].

USDA/USDI. 2009. [Internet site]. National land cover database 2001. Available at: http://www.mrlc.gov/nlcd.php. [Cited 20 April 2009].

Total of 28 references