Bridging the Gap Between Strategic and Management Forest Inventories

Ronald E. McRoberts[1]

Abstract.—Strategic forest inventory programs collect information for a large number of variables on a relatively sparse array of field plots. Data from these inventories are used to produce estimates for large areas such as states and provinces, regions, or countries. The purpose of management forest inventories is to guide management decisions for small areas such as stands. Management inventories collect data for a smaller number of variables using greater sampling intensities and produce estimates for small areas. Increasingly, management inventories are coming to be regarded as prohibitively expensive. A relatively inexpensive alternative is to use a combination of strategic inventory data and remotely sensed data such as satellite imagery to construct maps of forest attributes and then aggregate pixel predictions to obtain estimates of stand-level means. The study emphases included using forest inventory plot data, Landsat Thematic Mapper satellite imagery, and a modification of the k-Nearest Neighbors technique to construct stem density and basal area maps from which estimates of stand-level means were calculated. The results indicate that although unbiased and precise estimates of stand-level means for below canopy attributes may be difficult to obtain, stands may nevertheless be ranked quite accurately with respect to a combination of stem density and basal area. Such rankings inform management decisions by identifying stands that require immediate attention.

Introduction

The national forest inventories (NFI) of the countries of western and central Europe and North America may be characterized as strategic inventories. In the United States, the NFI is conducted by the Forest Inventory and Analysis (FIA) program of the Forest Service, U.S. Department of Agriculture. Strategic forest inventories typically feature observation or measurement of an array of plots distributed across entire countries or regions of countries according to a predetermined sampling design. Sampling intensities range from one plot per a few hundred hectares to one plot per 10,000s of hectares. Strategic inventory programs generally observe or measure a large number of variables for each plot and produce estimates for large areas such as states or provinces, regions, and countries. The objectives of management inventories, characterized as stand examinations in the United States, are to produce sufficiently accurate stand-level estimates to guide management decisions. Thus, management inventories focus on fewer variables for smaller stand-size areas but feature greater sampling intensities, perhaps as intense as multiple plots per hectare. Despite fewer variables, the required sampling intensities for management inventories result in considerable costs. In fact, in many countries management inventories are increasingly regarded as prohibitively expensive.

The emerging alternatives to ground plot-based management inventories require intensive use of fine resolution remotely sensed data (e.g., Flewelling 2009). The proposed uses of these data are to delineate individual tree crowns, either by border to border or for sample locations, and to produce accurate estimates of tree attributes such as crown dimensions, species, and height. These tree attribute estimates are then used as input to statistical models that are calibrated using ground plot data and used to predict tree diameters. Finally, the estimates of species, height, and diameter are then used to estimate individual tree volumes and biomasses, which are aggregated to produce

[1] Mathematical Statistician, U.S. Department of Agriculture, Forest Service, Northern Research Station, Forest Inventory and Analysis, 1992 Folwell Avenue, St. Paul, MN 55108. E-mail: rmcroberts@fs.fed.us.

estimates of stand-level means. These proposed alternative approaches are costly because fine resolution remotely sensed data and the ground plot data for calibrating the models must be acquired. In addition, these approaches are still mostly in research and development stages and have not yet been demonstrated to be operationally feasible.

Alternatives exploiting existing, accessible ground data and inexpensive satellite imagery have not been investigated. Possible approaches include using the k-Nearest Neighbors (k-NN) technique in combination with NFI data and 30-m x 30-m resolution Landsat satellite imagery. The k-NN technique is a nonparametric, multivariate approach to imputing values of variables observed or measured on sample units to units that are similar in a covariate space but without observations. The k-NN approach with inventory data and Landsat imagery has been shown to produce representative maps of forest attributes (McRoberts *et al.* 2002, Tomppo and Halme 2004). In addition, areal estimates of mean forest attributes obtained by averaging k-NN pixel predictions have been shown to be comparable to sample-based estimates for 10 km radius circular areas (McRoberts *et al.* 2007). These results suggest that aggregations of predictions over all stand pixels may produce estimates of stand-level means that are sufficiently accurate either to mitigate the necessity of conducting costly, labor-intensive, ground plot-based management inventories or to create efficiencies by informing decisions as to which stands require management inventories. The objective of the study was to estimate stand-level means of stem density (SD) (trees/hectare) and basal area (BA) (m^2/hectare) using the k-NN technique with FIA plot data and Landsat satellite imagery. Successfully achieving the objective would constitute at least a partial bridging of the gap between strategic and management inventories.

Data

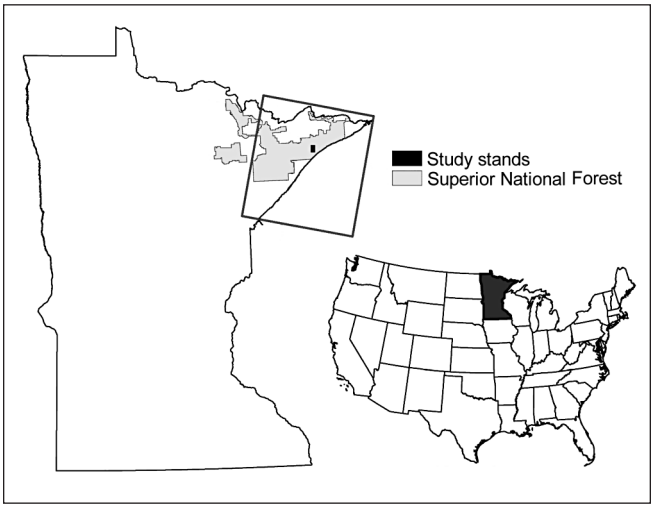
Satellite Image Data

The northern Minnesota study area was located wholly within the portion of the Superior National Forest in the path 26, row 27, Landsat scene (fig. 1). Three dates of Landsat Thematic Mapper (TM) imagery were obtained for this scene: November 1999 (fall), May 2003 (spring), and June 2003 (summer). Preliminary analyses indicated that the Normalized Difference Vegetation Index (NDVI) and the tassel cap (TC) transformations (brightness, greenness, and wetness) (Crist and Cicone 1984, Kauth and Thomas 1976) were superior to both the raw spectral band data and principal component transformations with respect to predicting forest attributes. Thus, NDVI and the three TC transformations for each of the three image dates were the 12 satellite image-based covariates.

FIA Plot Data

The FIA program has established field plot centers in permanent locations across the United States using a sampling design that is assumed to produce a random, equal probability sample (Bechtold and Patterson 2005, McRoberts *et al.* 2005). The plot array has been divided into five nonoverlapping, interpenetrating panels, and measurement of all plots in one panel is completed before measurement of plots in the next panel is initiated. Panels are selected on a 5-year rotating basis in the study area. Over a complete measurement cycle, the FIA sampling intensity is approximately one plot per 2,400 ha (6,000 acres). The State of Minnesota provides additional funding to double the sampling intensity to approximately one plot per 1,200 ha (3,000 acres). Each plot consists of four 7.31-m (24-ft) radius circular subplots. The subplots are configured as a central subplot and three peripheral subplots with centers located at 36.58 m (120 ft) and azimuths of 0, 120, and 240

Figure 1.—*Study area defined by the path 26, row 27 Landsat scene.*



degrees from the center of the central subplot. The path 26, row 27, Landsat scene included 1,086 FIA plots or 4,344 subplots measured between 1999 and 2003. For this study, observations and measurements on these plots were restricted to trees with diameter at breast height (d.b.h.) of 12.5 cm (5 in) or greater. Plot-level variables such as forest type, basal area per unit area, volume per unit area, and stem density per unit area may be calculated from the observed and measured variables.

In general, locations of forested or previously forested plots were determined using global positioning system receivers, while locations of nonforested plots were determined using aerial imagery and digitization methods. The spatial configuration of the FIA subplots with centers separated by 36.58 m and the 30-m x 30-m spatial resolution of the Landsat TM/ETM+ imagery permits individual subplots to be associated with individual image pixels. The subplot area of 167.87 m^2 (1,810 ft^2) is approximately 19 percent of the 900 m^2 (9,687 ft^2) pixel area, and subplot observations and measurements were assumed to characterize entire pixels.

Validation Data

Data from an independent study were used to assess the accuracy of the k-NN estimates of stand-level SD and BA means. Boundaries for 13 stands in the Superior National Forest study area, ranging in size from approximately 5 ha to approximately 27 ha, were delineated using segmentation techniques based on low altitude photography (table 1). In each stand, 6 to 15 temporary, variable radius plots were established using basal area factor 10; sampling intensities ranged from slightly more than one plot per hectare to more than two plots per hectare. Species were observed and d.b.h. was measured for all trees with d.b.h. greater than or equal to 12.5 cm (5 in).

Methods

k-NN Technique

The k-NN technique is an intuitive, nonparametric approach to either univariate or multivariate prediction based on the similarity in a feature space between the unit for which a prediction is desired and units for which observations are available. The set of 4,344 TM pixels for which corresponding subplot observations were available was designated the k-NN reference set, and the set of pixels with centers in the 13 stands was designated the k-NN target set. A basic implementation of the k-NN technique was used. The similarity measure in feature space was Euclidean distance,

$$d_{ij} = \sqrt{\sum_{l=1}^{L} \left(x_{il} - x_{jl} \right)^2}, \quad (1)$$

Table 1.—*Validation data.*

Stand	Area (ha)	No. plots	Sampling intensity (plots/ha)	SD		BA	
				Mean (trees/ha)	SE (trees/ha)	Mean (m^2/ha)	SE (m^2/ha)
1	18.5	8	2.3	122.3	52.4	46.3	17.6
2	7.1	7	1.0	160.1	30.2	25.7	4.8
3	5.3	6	0.9	222.9	51.4	108.3	8.7
4	10.4	12	8.6	137.2	31.2	70.8	8.5
5	10.2	9	1.1	189.4	32.1	113.3	14.1
6	5.9	7	0.8	257.6	98.8	104.3	24.6
7	4.8	6	0.8	256.7	25.3	41.7	4.0
8	17.1	8	2.1	269.4	39.0	117.5	25.5
9	27.4	15	1.8	171.0	24.1	117.3	14.0
10	16.9	9	1.9	197.9	47.2	68.9	9.5
11	10.5	11	1.0	240.1	42.6	89.1	7.4
12	9.1	9	1.0	594.8	42.6	128.9	7.4
13	10.9	9	1.2	242.0	42.3	65.6	9.7

BA = basal area. SD = stem density. SE = standard error.

where d_{ij} is the distance between the i^{th} target pixel and the j^{th} reference pixel, and l indexes the L spectral transformations, X, that define the feature space. The k-NN prediction for the m^{th} variable for the i^{th} target pixel is

$$\hat{y}_{i,m} = \frac{1}{k}\sum_{j=1}^{k} y_{j,m} \ , \qquad (2)$$

where j indexes the k nearest neighbor pixels in the reference set closest to the target pixel, m indexes the variables, and $y_{j,m}$ is the observation of the m^{th} variable for the subplot associated with the j^{th} nearest neighbor reference pixel.

Selection of Feature Space Variables

As noted by McRoberts *et al.* (2002), the inclusion in feature space of covariates unrelated to the dependent variables may be detrimental to the overall accuracy of predictions. Identification of the optimal set of feature space variables typically entails use of a bootstrapping technique in which the observations from a subset of the reference set are used to calculate predictions for the other subset. Accuracy is assessed by comparing the observations and predictions for the second subset. The leave-one-out technique is a special case of bootstrapping in which a prediction is calculated for each element of the reference set using the remaining reference set observations (Lachenbruch and Mickey 1986). Accuracy is assessed for continuous variables by comparing the observations and predictions for all reference set elements using root mean square error (RMSE) or a similar measure. When the reference set is large and the number of candidate feature space covariates is also large, an exhaustive evaluation of all covariate combinations can be extremely time-consuming.

An alternative to an exhaustive search of all covariate combinations is based on the genetic algorithm technique described by Tomppo (2004). Although this algorithm does not guarantee the optimal selection of covariates, it greatly reduces the selection time. A modification of this technique was used as follows:

1. Construct a random combination of covariates by independently selecting a random number, X, from a uniform (0, 1) distribution for each candidate feature space variable; if X > 0.5, include the covariate as a feature space variable in the combination; construct 100 such combinations.
2. Calculate RMSE for each of the 100 combinations using the leave-one-out technique.
3. Order the combinations by RMSE.
4. Select the 10 combinations with smallest RMSEs and construct 100 new combinations by pairwise crossing each combination with each other combination as follows:
    a. If a variable was included in both combinations of a pair, include it in the new combination.
    b. If a variable was not included in either combination of a pair, exclude it from the new combination.
    c. If a variable was included in one but not both combinations of a pair, randomly select a number, X, from a uniform (0, 1) distribution and include the variable in the new combination if X > 0.5.
    d. Following steps 4a through c, introduce random changes by randomly selecting a variable and reversing its inclusion or exclusion for the new combination; these changes are introduced only in new combinations constructed from pairs in which the first combination has greater RMSE from step 2 than the second combination.
5. Return to step 2.

Repeat steps 2 through 5 until the top 10 combinations determined in step 3 stabilize. Note that a pairwise crossing of a combination with itself in step 4 introduces no change and ensures that each of the top 10 combinations from the previous iteration enter into the next iteration.

The modified genetic algorithm was used to select covariate combinations for the feature space for SD and BA separately and in combination. When used to select combinations for SD and BA separately, the corresponding RMSE, denoted $RMSE_{min}$, was noted for each variable. The genetic algorithm was then used to determine a covariate combination for the two variables together. In step 2, RMSE was calculated separately for each variable for each combination. In step 3, for each variable, the ratio of the step 2 RMSE and its corresponding $RMSE_{min}$ were calculated and the average of the two ratios was used to order the combinations. Thus, the two variables, SD and BA, are weighted equally.

Stand-Level Estimation

Two approaches to k-NN stand-level estimation were investigated. The first approach entailed calculating estimates of SD and BA means for each stand by simply averaging the k-NN pixel predictions for all pixels with centers in the stand. For each pixel, the k-NN prediction for the m^{th} variable was calculated using (2), and for each stand the estimates of SD and BA means were calculated as

$$\overline{Y}_{kNN,m} = \frac{1}{N}\sum_{i=1}^{N}\hat{y}_{i,m} \ , \qquad (3)$$

where m denotes the variable, $\hat{y}_{i,m}$ is as defined in (2), i indexes pixels with centers in the stand, and N is the number of pixels in the stand.

The second approach exploits two assumptions that may contribute to improving k-NN pixel predictions. The first assumption is that stands exhibit internal homogeneity with respect to tree-level attributes such as species, diameter, and height, and stand-level attributes such as forest type, SD, and BA. The second assumption relates to the distribution of forest attributes for nearest neighbor reference pixels for a stand's target pixels. The relationship between forest attributes and the spectral values of satellite imagery is usually many-to-one in the sense that pixels with multiple forest conditions may be associated with similar sets of spectral values. Despite this many-to-one relationship, the second assumption is that most reference pixels selected as nearest neighbors for a target pixel have forest conditions similar to the actual conditions associated with that target pixel. These two assumptions may be used to modify the k-NN technique in four steps:

1. For each target pixel in a stand, select $m > k$ nearest neighbor reference pixels where k is the number intended to be used to calculate the k-NN predictions.
2. Determine the distribution of forest attributes for all m nearest neighbor reference pixels for all stand target pixels.
3. For each target pixel, create two suborderings of nearest neighbor reference pixels, one for pixels in the central portion of the step 2 distribution, and one for pixels in the tails of the distribution.
4. For each target pixel, modify the ordering of the nearest neighbor reference pixels by appending the pixels in the tails of the step 2 distribution to the end of ordering of pixels in the central portion of the distribution.

An example illustrates the modification. Suppose the stand-level SD distribution for all nearest neighbor reference pixels constructed in step 2 is as depicted in figure 2a. For a selected target pixel, neighbors 1, 2, and 4 of the original ordering of the $m = 5$ neighbors are in the central portion of the distribution, and neighbors 3 and 5 are in the tails of the distribution (fig. 2b). Thus, neighbors 3 and 5 are appended to the end of the ordering of neighbors 1, 2, and 4 to produce a modified ordering. For $k = 3$, the k-NN prediction for this pixel would then be based on original neighbors 1, 2, and 4 rather than on original neighbors 1, 2, and 3. Pixel-level predictions are calculated using (2) with the modified ordering of the nearest neighbors. Stand-level estimates are calculated using (3).

Figure 2a.—*Distribution of stem density for nearest neighbor pixels.*

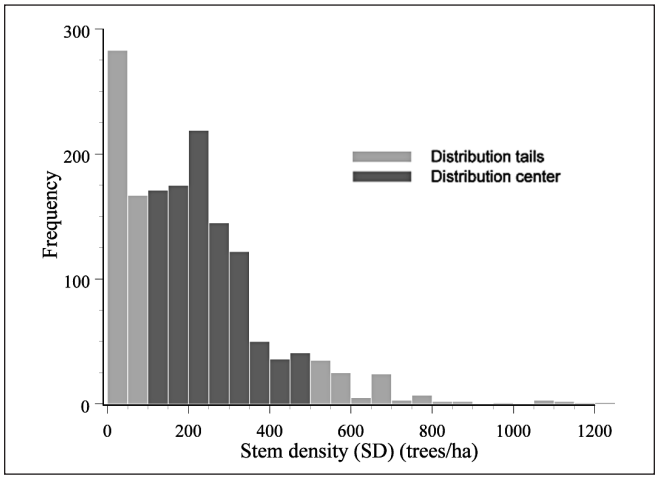
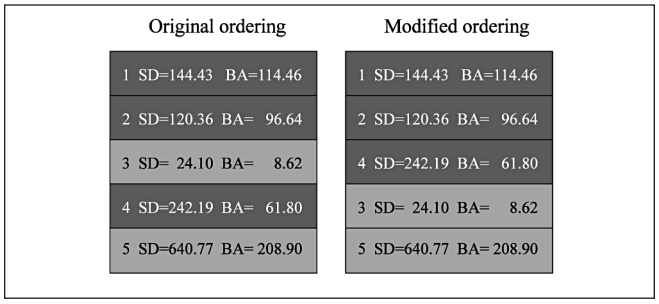


Figure 2b.—*Original and modified orderings of nearest neighbor reference pixels for a single target pixel.*



Validation Estimates

For each stand, the validation data were used to estimate stand-levels means as

$$\overline{Y}_{val,m} = \frac{1}{n}\sum_{i=1}^{n} y_{i,m}\ , \qquad (4)$$

where *val* indicates validation estimates, *m* designates the variable, *n* is the number of plots in the stand, and *i* indexes plots. The standard error of the mean was calculated as

$$SE\left(\bar{Y}_{val,m}\right) = \sqrt{\frac{\sum_{i=1}^{n}\left(y_{i,m} - \bar{Y}_{val,m}\right)^2}{n\left(n-1\right)}}. \qquad (5)$$

Comparisons

For each variable, estimates of the stand-level validation means were graphed against the corresponding original and modified k-NN estimates. Vertical lines extending 2-standard errors in both directions from the validation estimates were also graphed against the corresponding original and modified k-NN estimates. Under the hypothesis of no difference between the validation and k-NN stand-level estimates, the 2-standard error vertical lines should intersect the 1:1 line. As a method for comparing the validation and k-NN estimates, this graphic does not accommodate the uncertainty in the k-NN estimates and, therefore, may erroneously indicate statistically significant differences when none exists. Because the intent is to demonstrate that the k-NN estimates are not statistically significantly different than the validation estimates, this comparison may be characterized as conservative.

The preferred alternative to obtaining estimates of all stand-level means from ground plot observations is to produce remote sensing-based estimates that are sufficiently unbiased and precise to serve as the basis for stand management decisions. Even if the estimates are not sufficiently unbiased and precise for this purpose, however, they still may be useful for discriminating among stands with high likelihoods of requiring immediate treatment decisions or management inventories and those for which decisions may be delayed. If so, then the costly, labor-intensive, ground plot-based management inventories could be efficiently focused on the former class of stands.

A second approach to comparing the validation and k-NN estimates of the stand-level means was based on their utility for ordering stands with respect to the likelihood of requiring immediate management decisions. The assumption underlying this approach is that stands with a combination of large SD and large BA are more likely to require immediate decisions. For each variable, the 13 validation estimates of stand-level means and the two sets of 13 k-NN estimates were all divided by the largest validation estimate over the 13 stands. This calculation standardizes the estimates of the SD and BA means and permits them to be combined in a manner in which they are equally weighted. For each stand, the Euclidean distance from SD = 0 and BA = 0 of the standardized validation and k-NN estimates was calculated as

$$d = \sqrt{\overline{BA}_{std}^2 + \overline{SD}_{std}^2}\ , \qquad (6)$$

where the subscript *std* indicates standardized estimates. Stands with greater distances from SD = 0 and BA = 0 are interpreted as having greater combinations of SD and BA and, therefore, being in greater need of immediate management decisions. Orderings of stands based on this measure using the validation and the two sets of k-NN stand-level estimates were compared.

Results

For this study, the covariates selected for feature space using the genetic algorithm stabilized in five to eight iterations. Although some improvement in RMSE was achieved by selecting a subset of the 12 covariates for SD and BA separately and in combination, the improvement was not substantial (table 2). The five best combinations based on SD separately and in combination with BA had similar RMSEs; similarly, the five best combinations for BA separately and in combination with SD also had similar RMSEs. The RMSEs corresponding to the best covariate combinations selected when considering SD and BA simultaneously were nearly the same as for the RMSEs corresponding to combinations selected considering SD and BA separately. Both consistencies and inconsistencies were evident in the selection of covariates for feature space. Among the 12 spectral transformation covariates, summer brightness, spring

Table 2.—*Root mean square error.*

Variable combination rank	Separate		Combined		
	SD	BA	SD	BA	Proportion
1	182.1047	51.1821	182.3704	51.1935	1.0008
2	182.2339	51.2398	182.6407	51.2513	1.0021
3	182.6833	51.2444	182.2340	51.3858	1.0023
4	182.7614	51.2507	182.1047	51.4481	1.0026
5	182.8328	51.2438	182.8729	51.2538	1.0028
Best BA combination			184.0244	51.1821	1.0053
Best SD combination			182.1047	51.4481	1.0026
All 12 variables	184.3443	51.9538	184.3443	51.9538	1.0137

BA = basal area. SD = stem density.

and fall greenness, fall wetness, and summer and fall NDVI were selected in at least four of the five best combinations for SD, BA, and SD and BA simultaneously; spring NDVI was seldom selected for any of the five best combinations.

For the modified k-NN technique, the best results were obtained when using the distribution of SD based on the m = 5 nearest neighbors, when defining the central portion of the distribution to be between the 12th and 96th percentiles, and when using k = 3. Graphs of the validation estimates of the stand-level SD and BA means against the corresponding k-NN estimates for the 13 stands showed slightly better results when using the modified k-NN approach. For SD, 2-standard error confidence intervals for 12 of 13 stands cross the 1:1 line when using the modified k-NN approach, while 11 of 13 confidence intervals cross the line when using the original k-NN technique (fig. 3a). For BA, 2-standard error confidence intervals for 9 of 13 stands cross the 1:1 line when using the modified k-NN approach while 8 of 13 confidence intervals cross the line when using the original k-NN approach (fig. 3b).

The orderings of the stands obtained using the original and modified k-NN techniques were similar (table 3), although the modified technique produced slightly superior results relative to the ordering based on the validation data. An example from table 3 illustrates this analysis. The four highest ranked stands based on the largest estimates of SD and BA means from the validation data were stands 12, 8, 9, and 5; the four highest

Figure 3a.—*Validation versus modified k-Nearest Neighbor estimates of stand-level stem density means; each point represents a different stand; vertical lines are 95-percent confidence intervals.*

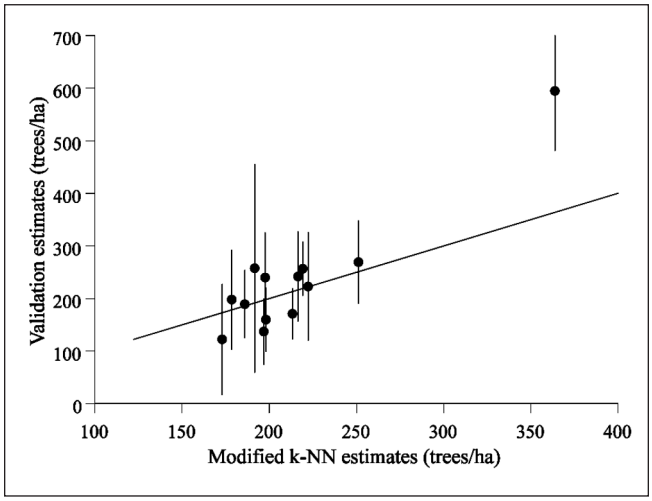


Figure 3b.—*Validation versus modified k-Nearest Neighbor estimates of stand-level basal area means; each point represents a different stand; vertical lines are 95-percent confidence intervals.*

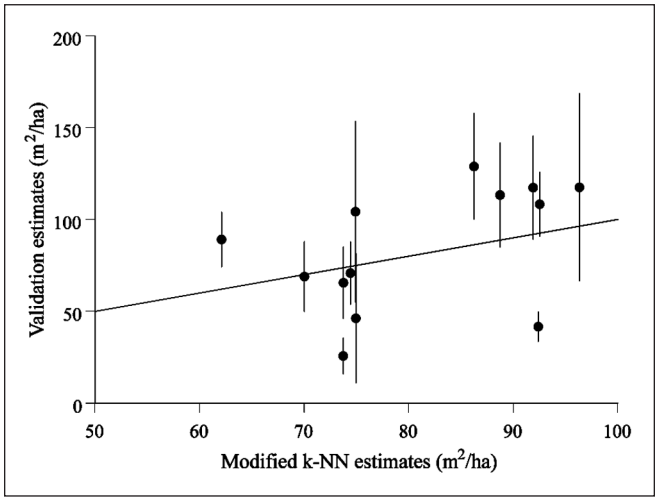


Table 3.—*Stand ranking statistics.*

Stand rank	Stand number (basis for ranking)			Cumulative proportion correctly ranked using k-NN estimates	
	Validation estimates	k-NN estimates			
		Original	Modified	Original	Modified
1	12	12	12	1.00	1.00
2	8	8	8	1.00	1.00
3	9	13	3	0.67	0.67
4	5	7	7	0.50	0.75
5	3	9	9	0.60	0.80
6	6	3	5	0.67	0.83
7	11	11	13	0.86	0.71
8	13	2	4	0.75	0.75
9	10	10	6	0.78	0.78
10	4	4	2	0.90	0.80
11	7	6	1	0.91	0.91
12	1	5	10	0.92	0.92
13	2	1	11	1.00	1.00

k-NN = k-Nearest Neighbor.

ranked stands using the original k-NN technique were stands 12, 8, 13, and 7; and the four highest ranked stands using the modified k-NN technique were stands 12, 8, 3, and 7. Thus, compared to the four highest ranked stands using the validation data, the modified k-NN technique correctly selected three stands for a proportion of 0.75, while the original k-NN technique correctly selected two stands for a proportion of 0.50.

Conclusions

Three conclusions may be drawn from this study. First, for these data, the genetic algorithm did not select covariates for the feature space that substantially decreased RMSEs. The similarity in RMSEs for the best covariate combinations when considering SD and BA separately and when considering them simultaneously can probably be at least partially attributed to this result. Second, the estimates of stand-level SD means obtained using the modified k-NN technique were nearly always within 95-percent confidence intervals of SD means obtained from the validation data. The estimates of stand-level BA means were less similar to the validation means, although the proportion within the confidence intervals was still great than 0.70. The somewhat lower proportion for BA can perhaps be attributed to two factors: (1) BA is essentially a below canopy attribute while optical sensors such as Landsat primarily respond to sunlight reflected from the forest canopy, and (2) the validation data were obtained from variable radius plots, while the data used to train the satellite imagery were obtained from FIA fixed radius plots. Because variable radius plots use BA as a factor for selecting trees to measure while fixed radius plots do not, estimates of BA obtained using these two different sampling approaches may not be entirely compatible. Third, the similarity in estimates of the stand-level SD and BA means obtained using the validation data and the means obtained using the TM imagery, the FIA plot data, and the k-NN technique suggest possibilities for effectively bridging the gap between strategic and management inventories. Even if the k-NN estimates of stand-level means are not deemed suitable to constitute the basis for stand management decisions, the high proportions of correctly ranked stands suggest that the k-NN estimates of means may be used to inform decisions as to which stands require immediate management inventories.

Future research should be conducted to extend this study in several directions. First, additional validation data should be obtained to expand the study beyond 13 stands. Second, additional studies should be conducted in other forest types. Third, validation data should be obtained using fixed radius plots of the size and configuration of FIA subplots.

Literature Cited

Bechtold, W.A.; Patterson, P.L., eds. 2005. The enhanced Forest Inventory and Analysis program—national sampling design and estimation procedures. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85 p.

Crist, E.P.; Cicone, R.C. 1984. Application of the tasseled cap concept to simulated Thematic Mapper data. Photogrammetric Engineering and Remote Sensing. 50: 343-352.

Flewelling, J. 2009. Forest inventory predictions from individual tree crowns: regression modeling within a sample framework. In: McRoberts, R.E.; Reams, G.A.; Van Deusen, P.A.; McWilliams, W.H., eds. Proceedings, eighth annual forest inventory and analysis symposium: 203-208.

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. Proceedings, symposium on machine processing of remotely sensed data. West Lafayette, IN: Purdue University: 41-51.

Lachenbruch, P.A.; Mickey, M.R. 1986. Estimation of error rates in discriminant analysis. Technometrics. 10: 1-11.

McRoberts, R.E.; Bechtold, W.A.; Patterson, P.L.; Scott, C.T.; Reams, G.A. 2005. The enhanced Forest Inventory and Analysis program of the USDA Forest Service: historical perspective and announcement of statistical documentation. Journal of Forestry. 103(6): 304-308.

McRoberts, R.E.; Nelson, M.D.; Wendt, D.G. 2002. Stratified estimation of forest area using satellite imagery, inventory data, and the k-nearest neighbors technique. Remote Sensing of Environment. 82: 457-468.

McRoberts, R.E.; Tomppo, E.O.; Finley, A.O.; Heikkinen, J. 2007. Estimating areal means and variances of forest attributes using the k-Nearest Neighbors technique and satellite imagery. Remote Sensing of Environment. 111: 466-480.

Tomppo, E.; Halme, M. 2004. Using coarse scale forest variables as ancillary information and weighting of information in k-NN estimation: a genetic algorithm. Remote Sensing of Environment. 92: 1-20.