Available online at www.sciencedirect.com
ScienceDirect
Remote Sensing of Environment 111 (2007) 466–480
Remote Sensing of Environment
www.elsevier.com/locate/rse

Estimating areal means and variances of forest attributes using the *k*-Nearest Neighbors technique and satellite imagery

Ronald E. McRoberts [a,*], Erkki O. Tomppo [b], Andrew O. Finley [c], Juha Heikkinen [b]

[a] *Northern Research Station, USDA Forest Service, St. Paul, MN, USA*
[b] *The Finnish Forest Research Institute, Helsinki, Finland*
[c] *Department of Forest Resources, University of Minnesota, St. Paul, MN, USA*

Received 2 October 2006; received in revised form 5 April 2007; accepted 7 April 2007

Abstract

The *k*-Nearest Neighbor (*k*-NN) technique has become extremely popular for a variety of forest inventory mapping and estimation applications. Much of this popularity may be attributed to the non-parametric, multivariate features of the technique, its intuitiveness, and its ease of use. When used with satellite imagery and forest inventory plot data, the technique has been shown to produce useful estimates of many forest attributes including forest/non-forest, volume, and basal area. However, variance estimators for quantifying the uncertainty of means or sums of *k*-NN pixel-level predictions for areas of interest (AOI) consisting of multiple pixels have not been reported. The primary objectives of the study were to derive variance estimators for AOI estimates obtained from *k*-NN predictions and to compare precision estimates resulting from different approaches to *k*-NN prediction and different interpretations of those predictions. The approaches were illustrated by estimating proportion forest area, tree volume per unit area, tree basal area per unit area, and tree density per unit area for 10-km AOIs. Estimates obtained using *k*-NN approaches and traditional inventory approaches were compared and found to be similar. Further, variance estimates based on different interpretations of *k*-NN predictions were similar. The results facilitate small area estimation and simultaneous and consistent mapping and estimation of multiple forest attributes.
Published by Elsevier Inc.

Keywords: Landsat; Probability-based inference; Model-based inference; Variance estimator

1. Introduction

Forest inventory programs typically report estimates for medium to large geographic areas using data collected from arrays of field plots. Due to budgetary constraints and natural variability in the resource, sufficient numbers of plots frequently cannot be observed to satisfy precision guidelines for the estimates of many variables unless the estimation process is enhanced using ancillary data. Satellite imagery has been accepted as a source of ancillary data that can be used with stratified estimation techniques to increase the precision of estimates with little corresponding increase in costs (Hansen & Wendt, 2000; Liknes et al., 2004; McRoberts et al., 2006, 2002a,b; Nilsson et al., 2005). In addition to greater estimation precision, users frequently request estimates for smaller areas than those reported and, ultimately, maps depicting the spatial distributions of forest resources. Although stratified estimation using classified satellite imagery can produce greater precision, it cannot address small area estimation and can only produce maps of broad categories of resources. The latter issues require more spatially intense sampling designs, more and different kinds of ancillary data, and/or methods that extract more information from the ancillary data. The increased costs associated with more intense sampling and larger suites of ancillary data often preclude these approaches. Therefore, approaches that make greater use of the satellite imagery merit consideration. One such approach, which is appropriate for mapping categorical forest attributes such as forest cover type, is to use the plot data to train classifiers of satellite imagery. A second approach, which is appropriate for mapping continuous attributes such as proportion forest area and tree volume, is to use the plot data to calibrate models that predict values of these forest attributes for individual pixels (McRoberts, 2006).

* Corresponding author. Tel.: +1 651 649 5174; fax: +1 651 649 5285.
E-mail address: mcroberts@fs.fed.us (R.E. McRoberts).

0034-4257/$ - see front matter. Published by Elsevier Inc.
doi:10.1016/j.rse.2007.04.002

Several aspects of the model-based approach merit consideration. First, the literature pertaining to variance estimation for model-based estimates for regions consisting of multiple pixels is sparse; Gregoire (1998), Aubry and Debouzie (2000), and McRoberts (2006) are among the few references. Second, unlike sample-based approaches, including stratified estimation, the issue of bias may be more difficult to assess for model-based approaches. Third, the highly correlated nature of suites of forest variables mitigates against separate, univariate models for individual variables. For example, univariate models for proportion forest area and volume may predict large volumes for pixels having small proportion forest area predictions and vice versa. Fourth, many multivariate approaches require that the response variables follow specified statistical distributions, usually Gaussian.

If issues associated with the model-based approach could be satisfactorily resolved, several advantages would accrue. First, in addition to estimates for medium and large regions, users would obtain spatial distributions of forest resources from which estimates for small areas could be calculated by aggregating individual pixel predictions. Second, the model-based approach has the potential to use more information from the satellite imagery than simply broad categories in the form of strata. Third, many natural resource inventory programs prohibit release of exact plot locations and proprietary plot information, thus limiting sample-based small area estimation and user-defined analyses. Estimates based on aggregations of model predictions for individual pixels circumvent this prohibition, because model predictions are typically aggregations of observations for multiple plots, and the particular plots, their locations, and their ownerships need not be revealed.

Among techniques that have been investigated for predicting forest attributes from satellite imagery and ground data, parametric regression (Ardö, 1992; Dungan, 1998; McRoberts, 2006; Tomppo, 1988) and the nonparametric *k*-Nearest Neighbor (*k*-NN) technique have been popular. The *k*-NN technique is an intuitive, non-parametric approach to either univariate or multivariate prediction based on the similarity in a covariate space between the unit for which a prediction of the response variable is desired and units for which observations of the response variable are available. The *k*-NN technique has gained popularity in forestry applications for mapping forest attributes using the spectral values of satellite imagery (Franco-Lopez et al., 2001; McRoberts et al., 2002a; Trotter et al., 1997). The maps have been used to depict the spatial distribution of forest attributes, to increase the precision of estimates via stratified estimation (McRoberts et al., 2002a; Nilsson et al., 2005), and to estimate the relative land area represented by individual plots (Tomppo & Halme, 2004). The *k*-NN technique has been particularly popular in Finland (Halme & Tomppo, 2001; Katila & Tomppo, 2001, 2002; Tomppo, 1991; Tomppo et al., 1999).

Because of its utility for mapping forest attributes and its lack of distributional assumptions, the *k*-NN technique was investigated for simultaneously mapping four forest attributes: proportion forest area (PFA), volume (m^3/ha) (VOL), basal area (m^2/ha) (BA), and stem density (tree count/ha) (D). Estimates for an area of interest (AOI) consisting of multiple pixels may be obtained by calculating sums or means for all pixels in the AOI. One disadvantage of this technique for inferential purposes, however, is that variance estimators for AOIs consisting of multiple pixels have not been reported. Thus, the objectives of the study were fourfold: (1) to construct compatible maps of PFA, VOL, BA, and D using the non-parametric, multivariate *k*-NN technique with forest inventory data and satellite imagery, (2) to compare estimates of PFA, VOL, BA, and D obtained by aggregating pixel predictions for multiple pixel AOIs and estimates obtained solely from inventory plot data, (3) to derive variance estimators for AOI estimates obtained from *k*-NN predictions, and (4) to compare precision estimates obtained when considering a *k*-NN prediction as an estimate of the mean of the distribution of possible observations for a particular covariate value or as an estimate of an individual observation from that distribution.

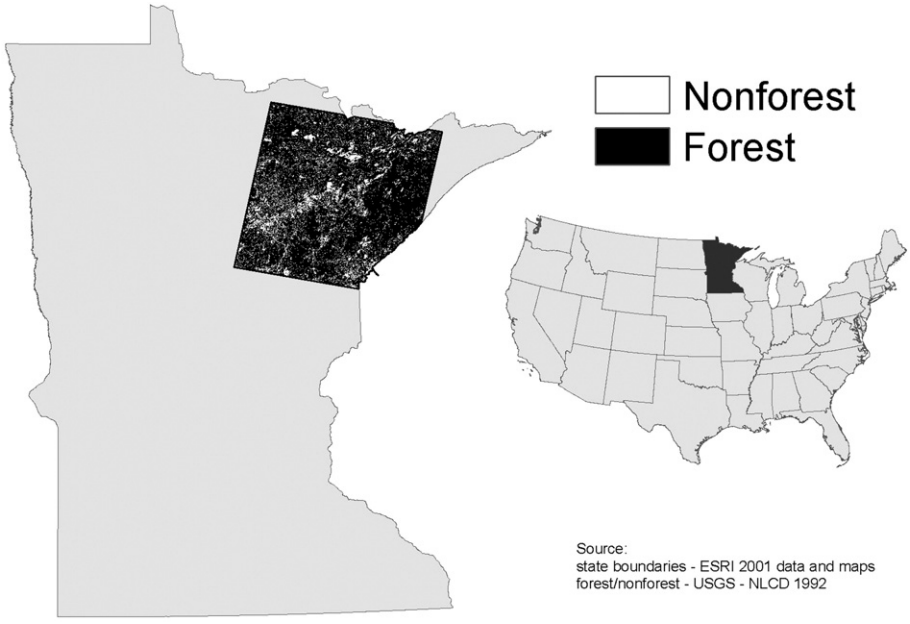

Fig. 1. Study area.

2. Data

2.1. Study area

The study area was located in northeastern Minnesota, USA, and was defined by Landsat scene row 27, path 27 (Fig. 1). Elevation in the study area is generally in the range 350–450 m, although small areas have elevations as great as 550 m. Land cover in the study area is approximately 75% forest land and is characterized by uneven-aged, naturally regenerated, mixtures of hardwoods and conifers dominated by aspen-birch and spruce-fir associations. Average daily temperatures range as low as −18 °C in winter and as high as 27 °C in summer. Average annual precipitation ranges from 60 cm to 75 cm, and average annual snowfall ranges from 115 cm to 255 cm.

2.2. Satellite imagery

Landsat imagery for row 27 of path 27, was obtained from the Multi-Resolution Land Characterization 2001 (MRLC 2001) land cover mapping project (Homer et al., 2004) of the U.S. Geological Survey (Fig. 1). The imagery was characterized by several salient features: (1) a combination of Landsat 5 TM and Landsat 7 ETM+ data, (2) geometrically and radiometrically corrected, (3) cubic convolution resampling to 30 m × 30 m spatial resolution, (4) visible and infrared bands (1–5,7), and (5) conversion to at-satellite reflectance. Imagery for three dates, April 2000, July 2001, and November 1999, corresponding to early, peak, and late vegetation green-up (Yang et al., 2001) were obtained for the scene. Preliminary analyses indicated that the Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1973) and the tassel cap (TC) transformations (brightness, greenness, and wetness) (Kauth & Thomas, 1976; Crist & Cicone, 1984) were superior to both the raw spectral band data and principal component transformations with respect to predicting the four forest attributes. Thus, 12 satellite image-based predictor variables were used, NDVI and the three TC transformations for each of the three image dates.

2.3. Forest inventory plot data

The Forest Inventory and Analysis (FIA) program of the USDA Forest Service has established field plots in permanent locations using a sampling design that is assumed to produce a random, equal probability sample (Bechtold & Patterson, 2005; McRoberts et al., 2005). The plot array has been divided into five non-overlapping, 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- or 10-year rotating basis, depending on the region of the country. Over a complete measurement cycle, the sampling intensity is approximately one plot per 2400 ha. Some states provide additional funding to double or triple the sample size in which case the sampling intensity is approximately one plot per 1200 ha or 800 ha, respectively. In general, locations of forested or previously forested plots are determined using global positioning system (GPS) receivers, while locations of non-forested plots are determined using aerial imagery and digitization methods. Each field plot consists of four 7.31-m radius circular subplots. The subplots are configured as a central subplot and three peripheral subplots with centers located at 36.58 m and azimuths of 0°, 120°, and 240° from the center of the central subplot. Field crews measure the diameter at breast height (DBH) (1.37 m) above bark and the height of each tree with DBH ≥ 12.5 cm. Statistical models are then used to predict the volume of each tree from the DBH and height measurements, and volumes of all trees with DBH ≥ 12.5 cm on each subplot are added to obtain an estimate of subplot volume per unit area. Field crews also obtain the proportions of subplot areas that satisfy specific land use class conditions. Subplot estimates of proportion forest area are then obtained by collapsing land use classes into forest and non-forest classes consistent with the FIA definition of forest land. For this study, all plots were observed between 1998 and 2003. Observations were available for 2266 plots consisting of 9064 subplots of which 2393 had no forest cover, 156 had partial forest cover, and 6515 were completely forested.

2.4. Combining forest inventory and satellite image data

The response variables observed or calculated for the FIA subplots were PFA, VOL, BA, and D, and the covariate space of independent variables was defined by the 12 satellite image-based variables, NDVI and the three TC transformations for each of the three dates. The spatial configuration of the FIA subplots with centers separated by 36.58 m and the 30-m × 30-m spatial resolution of the imagery permits individual subplots to be associated with individual image pixels. The set of all pixels was divided into a reference set consisting of pixels containing subplot centers and a target set consisting of pixels not containing subplot centers. The subplot area of 167.87 m^2 is approximately 19% of the 900 m^2 pixel area and is assumed to yield a representative sample of the four response variables for the pixel. Thus, for a pixel in the reference set, the subplot observations of PFA, VOL, BA, and D were attributed to the entire pixel.

3. Methods

Traditionally, inventory programs have used probability-based approaches to inference for estimates of forest attributes. Properties of variance estimators for these approaches are based on random variation resulting from the probabilities of selection of sampling units associated with the prescribed sampling design, thus their characterization as probability-based (Hansen, Madow, & Tepping, 1978). Although the term "design-based" is sometimes used to characterize these approaches, the term "design" in a sample survey context is not well-defined in the sense that it may refer simply to the selection of sample units or additionally to the entire inferential process (Kendall & Buckland, 1982). With probability-based approaches, the value of each response variable associated with a sampling unit is considered fixed. A crucial property of the combination of the FIA sampling design and probability-based estimators is that in expectation the sample mean is an asymptotically unbiased estimate of the population mean.

A second category of approaches to inference is characterized as model-based. With these approaches, the value of a variable associated with a sampling unit is considered a realization from an assumed probability distribution designated a superpopulation. The superpopulation parameters of interest include the mean, μ, and variance, σ^2, associated with the distribution of possible values for each sampling unit, and the correlation, ρ, among realizations between sampling units. With model-based approaches, both the estimates of means of distributions for the same values of covariates and the predictions of realizations from those distributions are based on the assumed model, and the assessment of their uncertainty is based on estimates of superpopulation variances. For inventory applications, areal estimates obtained by aggregating predictions for all sampling units in an AOI are often of interest.

Several distinctions between probability-based and model-based approaches merit noting. First, the population associated with a probability-based approach is only one of the possible populations that could be realized from the superpopulation, and the sample is only one of the possible samples that could be realized for the same sampling units. Second, inference with probability-based approaches requires estimates of population parameters, while inference with model-based approaches requires estimates of superpopulation parameters which are not even relevant with probability-based approaches. Third, unbiasedness of probability-based estimators depends only on assumptions concerning the sampling design, while unbiasedness of model-based estimators depends on the validity of the assumed superpopulation model. Fourth, variances obtained using probability-based approaches are generated from the sampling design, while variances obtained using model-based approaches result from random components in the assumed superpopulation model. Finally, finite population correction factors (Cochran, 1977) cause the limit of the probability-based variance of the estimate of a population mean as the sample approaches a complete census to be zero. Such would not be the case with model-based approaches.

Although estimates of forest attributes obtained using probability- and model-based approaches are based on different conceptual assumptions, at least two reasons for comparing them are relevant. First, the forest inventory community has traditionally used probability-based approaches, and the resulting estimates of population means have become a standard for comparison. Second, concern for model prediction bias would be alleviated if the estimate of the superpopulation mean obtained as the average of the estimates of means of distributions over all sampling units using a model-based approach were close to the estimate of the population mean obtained using a probability-based approach. Because the population assumed for probability-based approaches is only one of the possible populations that could be realized from the superpopulation, differences in estimates of the population mean and the superpopulation mean could be interpreted with respect to the distribution of possible population means. If the model is adequately formulated and the sample is representative of the population, then the estimate of the population mean should be close in a relative sense to the estimate of the superpopulation mean. If they are not close, then the model is biased, the probability-based sample is not representative of the population, or an unlikely population has been realized from the superpopulation. Attention is usually focused on the first two alternatives, although it is difficult to distinguish between them.

3.1. Probability-based approach

The FIA program uses a probability-based approach to inference based on a sampling design that is assumed to produce an equal probability sample (Bechtold & Patterson, 2005; McRoberts et al., 2005). The program attributes the aggregation of data for the four subplots of the FIA plot to the center point of the central subplot. The estimator of the population mean for an AOI is,

$$\bar{Y}_P = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{1}$$

where y_i is the observation of the response variable of interest for the ith plot, n is the sample size, and the subscript P denotes a probability-based estimator. Although the FIA program uses post-sampling stratification and stratified estimation, probability-based variance estimators for this study are based on an assumption of simple random sampling (SRS) because sufficient numbers of plots per stratum are not available for small AOIs which are of particular interest for this study. An estimate of the variance of $\bar{Y}_P$ is,

$$\hat{\mathrm{Var}}(\bar{Y}_P) = \frac{\sum_{i=1}^{n}(y_i - \bar{Y}_P)^2}{n(n-1)}. \tag{2}$$

When calculating $\bar{Y}_P$ for this study, data were included for all four subplots of each FIA plot with center in the AOI, regardless of whether parts of some subplots were outside the AOI. In addition, finite population correction factors (Cochran, 1977) were ignored because forest inventory sample sizes are orders of magnitude smaller than population sizes.

3.2. Model-based approach

Because the subplot observations of forest attributes are attributed to the entire Landsat pixel for this study, the pixel is hereafter considered the sampling unit even though the subplot area is slightly less than 20% of the pixel area.

3.2.1. The k-NN technique

The criteria for k-NN prediction include observations of the response variables for a subset of pixels and observations of covariates for all pixels. The set of pixels for which observations of both the response variables and the covariates are available is designated the reference set, and the set of all pixels for which predictions are sought is designated the target set. The k-NN prediction for the ith pixel is,

$$\tilde{y}_i = \left(\sum_{j=1}^{k} w_{ij}\right)^{-1}\left(\sum_{j=1}^{k} w_{ij} y_j^{\,i}\right) \tag{3}$$

where $\{y_j{}^i, j=1,..k\}$ is the set of observations for the k pixels in the reference set nearest to the ith pixel with respect to a distance metric, d, in the covariate space, and the set $\{w_{ij}, j=1,..k\}$ consists of pixel weights. Common selections for the distance metric are Mahalanobis distance (Kendall & Buckland, 1982), or weighted Euclidean distance,

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

where i and j denote arbitrary pixels, l indexes the covariates, x_{il} is the value of the lth covariate for the ith pixel, and the set $\{v_l\}$ consists of weights associated with individual covariates. Common selections for the pixel weights are $w_{ij}=1$, $w_{ij}=d_{ij}{}^{-1}$, or $w_{ij}=d_{ij}{}^{-2}$.

The k-NN technique is one of several nearest neighbor techniques that have gained popularity for natural resources applications. Other variations include the Most Similar Neighbor (MSN) technique for which $k=1$ and the distance metric is based on canonical correlation (Hassani et al., 2004; Moeur & Stage, 1995) and the Gradient Nearest Neighbor (GNN) technique for which $k=1$ and the distance metric is based on canonical correspondence (Ohmann & Gregory, 2002).

For this study, only the k-NN technique was considered. Selections of the k-value and the particular spectral variables used for determining distance in the covariate space are often obtained using a leave-one-out approach (Lachenbruch & Mickey, 1986) and a measure of the quality of predictions such as proportion correctly classified for categorical variables or root mean square error (RMS_e) for continuous variables. Frequently, however, all the available covariates are used in the distance measure, and an arbitrary but small k-value is selected. The rationale for selecting a small k-value is that when the k-NN technique is used to predict values for multiple variables simultaneously, predictions resulting from smaller k-values may better preserve observed covariance structures which results in more consistent estimates of multiple response variables.

The k-NN technique has attributes that make it appealing for predicting and mapping forest attributes. First, it can simultaneously impute multiple forest attributes to mapping units such as satellite image pixels, and second, it is nonparametric and does not require assumptions of Gaussian distributions or homogeneous variances among observations of response variables. However, because k-NN predictions may extrapolate poorly beyond the range of the data, the technique is less robust to violations of the assumption that the total range and variation of the covariates in the target set are represented in the reference set.

Katila and Tomppo (2001) provide an excellent survey of the early k-NN literature for estimating forest attributes. McRoberts et al. (2002a) provide an overview of the k-NN method, offer and illustrate several precautions that should be observed, and demonstrate how maps constructed with the k-NN technique may be used with stratified estimation to increase the precision of forest attribute estimates. Franco-Lopez et al. (2001) discuss the utility of the k-NN technique for mapping and estimating volume and forest type. For estimating volume at stand and regional scales, Mäkelä and Pekkarinen (2001) examined different pixel window and image segment approaches to Landsat TM feature extraction and later (Mäkelä & Pekkarinen, 2004) used Landsat TM and stand-level inventory data to predict species specific stand volume. Tokola et al. (1996) and Katila and Tomppo (2001) investigated the combined use of forest inventory plot data and satellite imagery to predict volume and other continuous variables. Holmström and Fransson (2003) explored a combination of SPOT-4 XS and SAR data for use with the k-NN technique.

Tomppo and Halme (2004), proposed an improved k-NN algorithm that uses large-scale variation of forest variables as ancillary data coupled with a genetic algorithm to derive optimal covariate weights. Fazakas and Nilsson (1996) and Tomppo et al. (2002) used Landsat TM and forest inventory plot data to produce k-NN predictions of forest volume. Fazakas & Nilsson calculated TM pixel averages for AVHRR pixels, while Tomppo et al. (2002) calculated averages for IRS-1C WiFS pixels. Both then used regression techniques to calibrate models of the relationships between the TM spectral averages and spectral values of the larger pixels, and both then applied the models to larger geographic areas.

3.2.2. k-NN estimation

Because a primary focus of the study was derivation of variance estimators rather than refinement of predictions, a basic approach to k-NN prediction was used; i.e., the unweighted Euclidean distance metric with pixel and covariate weights equal to a constant value of 1 was used. In the superpopulation context, let μ_i, and σ_i^2 denote the mean and variance, respectively, of the distribution of possible realizations of a response variable associated with the same vector, X_i, of covariate values as the ith pixel, and let y_i denote a realization from this distribution. The realization of a single response variable, y_i, may be expressed as,

$$y_i = \mu_i + \varepsilon_i \tag{4}$$

where $E(\varepsilon_i)=0$, $\mathrm{Var}(\varepsilon_i)=\sigma_i^2$, $\mathrm{Cov}(\varepsilon_i,\varepsilon_j)=\rho_{ij}\sigma_i\sigma_j$, and ρ denotes spatial correlation. Further, let the k-NN prediction be,

$$\tilde{y}_i = \frac{1}{k}\left(\sum_{j=1}^{k} y_j{}^i\right). \tag{5}$$

where the set $\{y_j{}^i: j=1,...,k\}$ consists of the observations for k pixels in the reference set nearest to the ith pixel in the covariate space with respect to the distance metric

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

and where l indexes the 12 spectral band transformations used as covariates.

3.2.3. Estimators for superpopulation parameters

Greater utility of the k-NN technique requires estimators for the parameters, μ_i, σ_i^2, and ρ_{ij}; estimators for the variances and covariances of estimates of μ_i and y_i; and estimators for the

variances of areal estimates obtained by aggregating pixel predictions for multiple pixel AOIs. As previously defined, let $\{y_j^{\,i}, j=1,..k\}$ denote the set of observations for the k pixels in the reference set that are nearest to the ith pixel with respect to the unweighted Euclidean distance metric. Assuming approximate symmetry in the distribution of nearest neighbors around X_i in the covariate space, and assuming $\mu_j^{\,i} \approx \mu_i$ where μ_i is the mean corresponding to X_i, and $\mu_j^{\,i}$ is the mean of the distribution of which $y_j^{\,i}$ is a realization, then $E(\tilde{y}_i) \approx \mu_i$. The derivation of variance estimators begins with the equivalence of the k-NN prediction, $\tilde{y}_i$, as the estimator of both the mean, μ_i and the realization, y_i; i.e., $\hat{\mu}_i = \hat{y}_i = \tilde{y}_i$. Draper and Smith (1981, pp. 28–30) articulate this equivalence in a regression context. They note that the regression model prediction for a given value of the covariate may be considered "the predicted mean value" of the response variable or "a predicted value of an individual observation" which will vary about the true mean value of the response variable.

Several approaches to developing an estimator for σ_i^2 for k-NN applications are feasible. The leave-one-out method (Lachenbruch & Mickey, 1986) or RMS_e method for which,

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (y_i - \tilde{y}_i)^2,$$

where i indexes the n pixels in the reference set is often used. However, this approach assumes homogeneity of variance, and, as Kim and Tomppo (2006) note, it does not produce a direct measure of uncertainty for k-NN predictions and does not accommodate dependence due to spectral similarity of neighboring pixels or spatial dependence among observations. Kim and Tomppo (2006) propose an approach that uses a model of the covariate space variogram. Although this approach merits consideration, variance estimators for aggregations of pixel predictions for multiple pixel AOIs have not been derived. A third approach developed for this study incorporates spatial correlation and features an estimator of σ_i^2 as,

$$\hat{\sigma}_i^2 = \frac{\sum_{k=1}^{k} (y_j^{\,i} - \tilde{y}_i)^2}{k - \frac{1}{k}\sum_{j_1=1}^{k}\sum_{j_2=1}^{k} \rho_{j_1 j_2}}, \tag{6a}$$

(Appendix A.1) where j, j_1 and j_2 index the k neighbors in covariate space nearest to the ith pixel, and $\rho_{j_1 j_2}$ is the spatial correlation between the observations for the j_1th and the j_2th nearest neighbors. In the absence of spatial correlation, $\hat{\sigma}_i^2$ from (6a) reduces to,

$$\hat{\sigma}_i^2 = \frac{\sum_{j=1}^{k} (y_j^{\,i} - \tilde{y}_i)^2}{k-1}. \tag{6b}$$

Because of fragmentation in the predominantly forested landscapes resulting from features such as lakes, roads, and clearings, spatial correlation is not continuous; nevertheless, an assumption of continuous spatial correlation leads to a useful approximation. Under this continuity assumption, an estimator for ρ may be obtained using the variogram approach of Gumpertz et al. (2000):

(1) calculate the standardized residuals,

$$\hat{\delta}_i = \frac{y_i - \tilde{y}_i}{\hat{\sigma}_i};$$

(2) construct an empirical semi-variogram,

$$\hat{\gamma}(d) = \frac{1}{2||N(d)||}\sum_{N(d)} \left(\hat{\delta}_i - \hat{\delta}_j\right)^2$$

where $N(d)$ denotes a collection of pairs, $(\hat{\delta}_i, \hat{\delta}_j)$, whose Euclidean distance apart, d_{ij}, in geographic space lies within a given neighborhood of d, and $||N(d)||$ denotes the number of pairs in $N(d)$.

(3) fit a semivariogram (e.g., the exponential model),

$$\hat{\gamma}(h) = \hat{\alpha}_0 + \hat{\alpha}_1[1 - \exp(\hat{\alpha}_2 h)] \tag{7a}$$

to the empirical semi-variogram;

(4) and calculate,

$$\hat{\rho}_{ij} = 1 - \frac{\hat{\gamma}(d_{ij})}{\hat{\gamma}_{\text{total}}} = 1 - \frac{\hat{\alpha}_0 + \hat{\alpha}_1[1 - \exp(\hat{\alpha}_2 d_{ij})]}{\hat{\alpha}_0 + \hat{\alpha}_1}. \tag{7b}$$

An approach that iterates between estimating σ_i^2 from (6a) and the steps leading to (7a,b) should be used; convergence may be expected after only a few iterations. For the first iteration, σ_i^2 may be estimated from (6b) under the assumption of no spatial correlation.

3.2.4. Variance estimators

The covariance between estimates of the ith and jth means, $\hat{\mu}_i$ and $\hat{\mu}_j$, may be approximated as,

$$\text{Cov}(\hat{\mu}_i, \hat{\mu}_j) \approx \frac{\sigma_i \sigma_j}{k^2}\sum_{l_i=1}^{k}\sum_{l_j=1}^{k} \rho_{l_i l_j}, \tag{8a}$$

(Appendix A.2) where l_i and l_j index the k neighbors nearest to the ith and jth pixels, respectively. In the absence of spatial correlation, $\text{Cov}(\hat{\mu}_i, \hat{\mu}_j)$ from (8a) reduces to,

$$\text{Cov}(\hat{\mu}_i, \hat{\mu}_j) \approx \frac{m_{ij}\sigma_i \sigma_j}{k^2}, \tag{8b}$$

where m_{ij} is number of nearest neighbors common to both the ith and jth pixels. In the absence of spatial correlation, and if the ith and jth pixels share no common nearest neighbors, $\text{Cov}(\hat{\mu}_i, \hat{\mu}_j) = 0$.

The estimate of the variance of the ith mean, $\hat{\mu}_i$, is simply the special case of (8a) where $i=j$, so that,

$$\text{Var}(\hat{\mu}_i) \approx \frac{\sigma_i^2}{k^2}\sum_{l_1=1}^{k}\sum_{l_2=1}^{k} \rho_{l_1 l_2} \tag{9a}$$

where l_1 and l_2 index the k neighbors nearest to the ith pixel. In the absence of spatial correlation, (9a) reduces to,

$$\text{Var}(\hat{\mu}_i) \approx \frac{\sigma_i^2}{k}. \tag{9b}$$

The covariance between predictions of realizations, y_i and y_j, could be formulated in two ways: either $\mathrm{Cov}(\hat{y}_i, \hat{y}_j) = E[(\hat{y}_i - y_i)(\hat{y}_j - y_j)]$ or $\mathrm{Cov}(\hat{y}_i, \hat{y}_j) = E[(\hat{y}_i - \mu_i)(\hat{y}_j - \mu_j)]$. Because the emphasis is on predictions of realizations, the first formulation is used for this study. An estimator may then be derived as,

$$\mathrm{Cov}\left(\hat{y}_i, \hat{y}_j\right) \approx \frac{\sigma_i \sigma_j}{k^2}\left(\sum_{l_i=1}^{k}\sum_{l_j=1}^{k} \rho_{l_i l_j} - k\sum_{l_i=1}^{k} \rho_{l_i j} - k\sum_{l_j=1}^{k} \rho_{i l_j} + k^2\right), \tag{10a}$$

(Appendix A.3) where l_i and l_j index the k neighbors nearest to the ith and jth pixels, respectively. In the absence of spatial correlation, and if the ith and jth pixels are both included in the reference set, then for exactly one l_i and one l_j, $\rho_{l_i j} = \rho_{i l_j} = 1$. In this case, (10a) reduces to,

$$\mathrm{Cov}\left(\hat{y}_i, \hat{y}_j\right) \approx \sigma_i \sigma_j\left(1 - \frac{2}{k} + \frac{m_{ij}}{k^2}\right) \tag{10b}$$

where m_{ij} is the number of nearest neighbors common to both the ith and jth pixels. If only one of the ith or jth pixels is included in the reference set, then in the absence of spatial correlation, (10a) reduces to,

$$\mathrm{Cov}\left(\hat{y}_i, \hat{y}_j\right) \approx \sigma_i \sigma_j\left(1 - \frac{1}{k} + \frac{m_{ij}}{k^2}\right), \tag{10c}$$

and if neither the ith or jth pixels is included in the reference set, then in the absence of spatial correlation, (10a) reduces to,

$$\mathrm{Cov}\left(\hat{y}_i, \hat{y}_j\right) \approx \sigma_i \sigma_j\left(1 + \frac{m_{ij}}{k^2}\right). \tag{10d}$$

In the absence of spatial correlation, and if the ith and jth pixels share no common nearest neighbors, $\mathrm{Cov}(\hat{y}_i, \hat{y}_j) = \sigma_i \sigma_j$.

The estimate of the variance of the prediction of a realization, y_i, is simply the special case of (10a) where $i = j$, so that,

$$\mathrm{Var}(\hat{y}_i) \approx \frac{\sigma_i^2}{k^2}\left(\sum_{l_1=1}^{k}\sum_{l_2=1}^{k} \rho_{l_1 l_2} - 2k\sum_{l=1}^{k} \rho_{l i} + k^2\right) \tag{11a}$$

where l, l_1, and l_2 index the k neighbors nearest to the ith pixel. In the absence of spatial correlation, and if the ith pixel is included in the reference set, then (11a) reduces to,

$$\mathrm{Var}(\hat{y}_i) \approx \sigma_i^2\left(1 - \frac{1}{k}\right), \tag{11b}$$

otherwise if the ith pixel is not included in the reference set, then in the absence of spatial correlation, (11a) reduces to

$$\mathrm{Var}(\hat{y}_i) \approx \sigma_i^2\left(1 + \frac{1}{k}\right). \tag{11c}$$

3.2.5. Estimation for aggregations of pixel predictions

Two approaches may be considered for estimating the mean of a response variable for an AOI consisting of multiple pixels. First, the estimate may be obtained as the mean over estimates of the means, μ_i,

$$\overline{Y}_{M1} = \frac{1}{N}\sum_{i=1}^{N} \hat{\mu}_i, \tag{12}$$

where i indexes the N pixels in the AOI, $\hat{\mu}_i = \tilde{y}_i$ from (5), and the subscript $M1$ denotes the first of two estimators. An estimate obtained using (12) may be interpreted as the mean over all possible realizations of populations from the superpopulation. Second, the estimate may be obtained as the mean over predictions of realizations from the superpopulation,

$$\overline{Y}_{M2} = \frac{1}{N}\sum_{i=1}^{N} \hat{y}_i, \tag{13}$$

where $\hat{y}_i = \tilde{y}_i$ and $M2$ denotes the second estimator. An estimate obtained using (13) may be interpreted as the mean over one realization of a population from the superpopulation. Because $\hat{\mu}_i = \hat{y}_i = \tilde{y}_i$, then $\overline{Y}_{M1} = \overline{Y}_{M2}$ and $E(\overline{Y}_{M1}) = E(\overline{Y}_{M2}) = \frac{1}{N}\sum_{i=1}^{N} \mu_i$.

The variance of $\overline{Y}_{M1}$ is,

$$\mathrm{Var}(\overline{Y}_{M1}) = \frac{1}{N^2}\mathrm{Var}\left(\sum_{i=1}^{N} \hat{\mu}_i\right) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Cov}\left(\hat{\mu}_i, \hat{\mu}_j\right). \tag{14a}$$

Substituting from (8a) into (14a),

$$\mathrm{Var}(\overline{Y}_{M1}) \approx \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\frac{\sigma_i \sigma_j}{k^2}\left(\sum_{l_i=1}^{k}\sum_{l_j=1}^{k} \rho_{l_i l_j}\right)\right], \tag{14b}$$

where l_i and l_j index the k neighbors nearest to the ith and jth pixels, respectively. In the absence of spatial correlation, $\mathrm{Var}(\overline{Y}_{M1})$ from (14b) reduces to,

$$\mathrm{Var}(\overline{Y}_{M1}) \approx \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \sigma_i \sigma_j \frac{m_{ij}}{k^2}, \tag{14c}$$

where m_{ij} is the number of nearest neighbors common to both the ith and jth pixels.

The variance of $\overline{Y}_{M2}$ is,

$$\mathrm{Var}(\overline{Y}_{M2}) = \frac{1}{N^2}\mathrm{Var}\left(\sum_{i=1}^{N} \hat{y}_i\right) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Cov}\left(\hat{y}_i, \hat{y}_j\right). \tag{15a}$$

Substituting from (10a) into (15a),

$$\mathrm{Var}(\overline{Y}_{M2}) \approx \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\frac{\sigma_i \sigma_j}{k^2}\left(\sum_{l_i=1}^{k}\sum_{l_j=1}^{k} \rho_{l_i l_j} - k\sum_{l_i=1}^{k} \rho_{l_i j} - k\sum_{l_j=1}^{k} \rho_{i l_j} + k^2\right)\right], \tag{15b}$$

and, after substitution from (14b),

$$\mathrm{Var}(\bar{Y}_{M2}) \approx \mathrm{Var}(\bar{Y}_{M1}) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\frac{\sigma_i\sigma_j}{k^2}\left(k^2 - k\sum_{l_i=1}^{k}\rho_{l_ij} - k\sum_{l_j=1}^{k}\rho_{il_j}\right)\right], \quad (15c)$$

where l_i and l_j index the k neighbors nearest to the ith and jth pixels, respectively. Thus, $E(\bar{Y}_{M1}) = E(\bar{Y}_{M2})$, but $\mathrm{Var}(\bar{Y}_{M1}) \neq \mathrm{Var}(\bar{Y}_{M2})$. When considering $\mathrm{Var}(\bar{Y}_{M2})$ in the absence of spatial correlation, the effects of the ith or jth pixels being in the reference set are ignored, because the target set is assumed to be orders of magnitude larger than the reference set. Thus, in the absence of spatial correlation, (15b) reduces to,

$$\mathrm{Var}(\bar{Y}_{M2}) \approx \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\sigma_i\sigma_j\left(1 + \frac{m_{ij}}{k^2}\right)\right], \quad (15d)$$

where m_{ij} is the number of nearest neighbors common to both the ith and jth pixels.

All covariances and variances, $\mathrm{Cov}(\hat{\mu}_i, \hat{\mu}_j)$, $\mathrm{Var}(\hat{\mu}_i)$, $\mathrm{Cov}(\hat{y}_i, \hat{y}_j)$, $\mathrm{Var}(\hat{y}_i)$, $\mathrm{Var}(\bar{Y}_{M1})$, and $\mathrm{Var}(\bar{Y}_{M2})$ are now expressed in terms of only the variances, σ_i^2, and the correlations, ρ_{ij}. Substituting the estimator for σ_i^2 from (6a,b) and the estimator for ρ_{ij} from (7a,b) into (8a, b), (9a,b), (10a,b,c,d), (11a,b,c), (14a,b,c), and (15a,b,c,d) provides estimators for the variances and covariances of means and individual predictions and for the variances of areal estimates obtained as aggregations of pixel predictions for multiple pixel AOIs.

3.3. Analyses

Although the primary emphases of the study were related to issues of precision rather than bias, derivation of the variance estimators assumed unbiasedness in pixel predictions. Thus, investigation of the validity of this assumption was investigated using the reference set observations. First, the leave-one-out technique was used to obtain k-NN predictions for each reference set pixel with the constraint that only one nearest neighbor could be selected from among pixels corresponding to subplots of the same FIA plot. Second, for each response variable, the residuals resulting from these predictions were graphed against four measures: (1) observations, (2) standard deviations among observations for the five nearest neighbors, (3) feature space distances to the fifth nearest neighbors, and (4) distances to feature space center. The latter measure addresses the concern that at the edges of feature space, sparseness in the distribution of reference set pixels may lead to bias as the result of greater feature space distances to nearest neighbors and greater standard deviations among those observations for those nearest neighbors.

For each of the four response variables, two sets of graphs were constructed. First, the residuals were graphed against the four measures. This set of graphs indicates the general pattern of the distribution of residuals for each measure and provides information on heterogeneity of variance of the residuals. However, because of the large number of reference set observations, determining if the residuals were generally symmetrically distributed around zero was difficult. Therefore, second, the residuals were ordered with respect to the values of each measure and divided into equal size groupings beginning with the smallest value of the measure. The means of the residuals and the measures were calculated for each grouping, and the residual mean was graphed against the measure mean. This set of graphs indicates whether there are systematic patterns of deviations of residuals from a mean residual of zero.

The association of each of the four subplots of the FIA plot with a different pixel raises issues of the effects of spatial correlation. The approximately 35–40 m separation among subplots is considerably less than the expected effective ranges of spatial correlation among residuals. Therefore, because of the expected similarity among observations for subplots of the same plot and, hence, the expected similarity in spectral values for the associated pixels, the nearest neighbors for a particular pixel may tend to include pixels associated with multiple subplots of the same plot. This phenomenon may have two effects on variance estimates. First, the similarity among the assumed spatially correlated observations for subplots of the same plot may decrease the denominator in (6a), thus increasing the estimate of σ_i^2. Second, this same spatial correlation may also decrease the numerator of (6a), thus decreasing the estimate of σ_i^2. To assess these effects, two k-NN predictions were calculated for each pixel: the first permitted only one nearest neighbor from among the pixels associated with the four subplots of each FIA plot, and the second permitted multiple neighbors.

Within the study area, 15 points were selected to serve as centers of 10-km radius AOIs. The AOI centers were systematically distributed throughout the study area with minor adjustment to ensure they represented the range of conditions in the scene. For each pixel in each AOI, k-NN predictions for PFA, VOL, BA, and D were calculated for each combination of $1 \le k \le 10$ and for both single and multiple nearest neighbors associated with subplots of the same FIA plot. For each AOI and each combination, $\bar{Y}_P$ and $\bar{Y}_M = \bar{Y}_{M1} = \bar{Y}_{M2}$ were calculated. For each AOI and number of permitted neighbors among subplots of the same FIA plot, $\hat{\mathrm{Var}}(\bar{Y}_{M1})$ and $\hat{\mathrm{Var}}(\bar{Y}_{M2})$ were calculated for the selected k-value. Because neither $\bar{Y}_P$ nor $\bar{Y}_M$ are without uncertainty, and because they are correlated as a result of their basis in the same underlying data, no rigorous statistical test for comparing them was used. Rather, $\bar{Y}_P$ and $\bar{Y}_M$ were compared graphically and both root mean square and mean absolute differences were calculated.

Although the selection of k could be based on comparisons of pixel-level observations and predictions as is often done, this approach tends to produce large pixel-level RMSEs which do not, however, necessarily translate to large variances for areal estimates, a criterion of interest for this study. Therefore, the selection of k was based on comparisons of $\bar{Y}_P$ and $\bar{Y}_M$ for the 15 AOIs. Two approaches were used. First, for $1 \le k \le 10$ and for each combination of forest attribute variable and AOI,

$$\tau = \frac{\bar{Y}_P - \bar{Y}_M}{SE(\bar{Y}_P)} \quad (16)$$

was calculated, and then, for each variable and k-value, the mean of τ over the 15 AOIs was calculated. Although τ is similar to Student's t-statistic, it does not incorporate the uncertainty in $\bar{Y}_M$, and it does not incorporate the covariance between $\bar{Y}_P$ and

Table 1

Area of interest	Probability-based approach			Model-based approach [a]					
	No. plots	$\overline{Y}_P$	$\sqrt{\hat{\mathrm{Var}}(\overline{Y}_P)}$	1 neighbor only per plot			Multiple neighbors per plot		
				$\overline{Y}_M$	$\sqrt{\hat{\mathrm{Var}}(\overline{Y}_{M1})}$	$\sqrt{\hat{\mathrm{Var}}(\overline{Y}_{M2})}$	$\overline{Y}_M$	$\sqrt{\hat{\mathrm{Var}}(\overline{Y}_{M1})}$	$\sqrt{\hat{\mathrm{Var}}(\overline{Y}_{M2})}$
a. Proportion forest area estimates									
1	24	0.8607	0.0560	0.8594	0.00523	0.00529	0.8601	0.00519	0.00523
2	25	0.7866	0.0642	0.8182	0.00470	0.00479	0.8205	0.00446	0.00453
3	23	0.7935	0.0697	0.7365	0.00516	0.00522	0.7356	0.00493	0.00497
4	21	0.8929	0.0407	0.8460	0.00664	0.00666	0.8460	0.00632	0.00634
5	20	0.9625	0.0205	0.9075	0.00691	0.00696	0.9037	0.00692	0.00695
6	25	0.6684	0.0772	0.6604	0.00657	0.00667	0.6554	0.00590	0.00598
7	22	0.7169	0.0659	0.8307	0.00481	0.00492	0.8302	0.00460	0.00468
8	24	0.8120	0.0501	0.8101	0.00494	0.00504	0.8130	0.00458	0.00466
9	24	0.6797	0.0751	0.6559	0.00623	0.00635	0.6523	0.00588	0.00596
10	19	0.6556	0.0746	0.6963	0.00709	0.00719	0.6883	0.00662	0.00670
11	22	0.6538	0.0727	0.6969	0.00716	0.00728	0.6986	0.00686	0.00694
12	23	0.7717	0.0701	0.8455	0.00660	0.00671	0.7887	0.00625	0.00633
13	24	0.6146	0.0752	0.6427	0.00648	0.00661	0.6389	0.00620	0.00629
14	19	0.8491	0.0546	0.8758	0.00636	0.00641	0.8734	0.00581	0.00585
15	24	0.6792	0.0730	0.7956	0.00560	0.00572	0.7910	0.00521	0.00531
b. Volume (m^3/ha) estimates									
1	24	75.50	10.78	56.76	1.66	1.67	56.84	1.76	1.77
2	25	56.08	10.33	51.98	1.19	1.21	51.98	1.15	1.17
3	23	57.82	10.50	56.60	1.48	1.49	57.45	1.51	1.52
4	21	67.48	9.42	64.89	2.31	2.32	65.53	2.27	2.27
5	20	81.43	13.51	74.59	2.50	2.51	73.05	2.41	2.42
6	25	37.77	9.09	40.67	1.33	1.34	39.94	1.30	1.31
7	22	34.28	7.94	53.57	1.18	1.19	54.31	1.15	1.16
8	24	52.88	6.45	49.12	1.20	1.22	48.83	1.17	1.18
9	24	61.98	10.85	41.53	1.23	1.25	42.55	1.23	1.25
10	19	37.23	6.29	50.11	1.64	1.65	50.85	1.52	1.53
11	22	31.45	8.16	41.33	1.33	1.34	41.44	1.33	1.34
12	23	55.32	8.15	51.80	1.57	1.58	52.55	1.57	1.59
13	24	34.99	6.93	39.75	1.20	1.21	39.80	1.14	1.15
14	19	75.55	9.18	74.62	2.14	2.15	74.92	2.08	2.09
15	24	35.33	7.81	49.68	1.14	1.15	49.13	1.10	1.12
c. Basal area per acre (m^2/ha) estimates									
1	24	14.860	2.064	10.637	0.283	0.285	10.715	0.287	0.289
2	25	9.634	1.681	9.430	0.195	0.197	9.411	0.189	0.189
3	23	9.912	1.655	10.300	0.245	0.247	10.360	0.243	0.245
4	21	12.674	1.476	11.942	0.387	0.388	12.033	0.373	0.374
5	20	13.432	1.942	13.827	0.410	0.412	13.592	0.388	0.390
6	25	7.168	1.683	7.338	0.211	0.213	7.219	0.208	0.210
7	22	6.681	1.368	9.689	0.196	0.199	9.770	0.189	0.192
8	24	10.344	1.290	8.975	0.198	0.189	8.943	0.192	0.194
9	24	10.523	1.690	7.276	0.202	0.204	7.421	0.199	0.201
10	19	6.785	1.068	8.736	0.249	0.251	8.821	0.236	0.238
11	22	5.602	1.380	7.291	0.206	0.208	7.285	0.208	0.210
12	23	10.346	1.449	9.251	0.246	0.249	9.372	0.253	0.255
13	24	6.117	1.102	6.998	0.193	0.195	6.989	0.188	0.190
14	19	12.678	1.433	13.537	0.343	0.345	13.576	0.335	0.337
15	24	6.789	1.582	8.906	0.186	0.188	8.808	0.450	0.455

$\overline{Y}_M$ as a result of both being based on the same sample data. Second, a multivariate approach was used in which,

$$\psi = (\overline{Y}_P^* - \overline{Y}_M^*)' V^{-1} (\overline{Y}_P^* - \overline{Y}_M^*) \quad (17)$$

was calculated for $1 \le k \le 10$ and each AOI where the superscript, *, indicates a vector of means for all four response variables, and V is the 4×4 matrix of covariances for the plot observations of the four variables for the AOI. For each k-value, the mean of ψ was then calculated over the 15 AOIs. Although ψ is similar to a χ^2 statistic (Graybill, 1961), it does not incorporate the uncertainty in $\overline{Y}_M^*$ or the covariances between $\overline{Y}_P^*$ and $\overline{Y}_M^*$ as a result of both being based on the same sample data.

For the selected k-value, each response variable, and each permitted number of neighbors associated with subplots of the same FIA plot, spatial variability among the standardized residuals was evaluated for the entire study area under the assumptions of

Table 1 (*continued*)

Area of interest	Probability-based approach			Model-based approach[a]					
	No. plots	$\overline{Y}_P$	$\sqrt{\hat{\mathrm{Var}}(\overline{Y}_P)}$	1 neighbor only per plot			Multiple neighbors per plot		
				$\overline{Y}_M$	$\sqrt{\hat{\mathrm{Var}}(\overline{Y}_{M1})}$	$\sqrt{\hat{\mathrm{Var}}(\overline{Y}_{M2})}$	$\overline{Y}_M$	$\sqrt{\hat{\mathrm{Var}}(\overline{Y}_{M1})}$	$\sqrt{\hat{\mathrm{Var}}(\overline{Y}_{M2})}$
d. Stem density estimates									
1	24	437.24	60.96	318.44	7.81	7.87	321.43	7.57	7.63
2	25	257.01	46.13	267.31	5.07	5.13	265.31	4.92	4.98
3	23	245.10	41.09	292.47	6.57	6.63	289.16	6.25	6.30
4	21	404.40	34.35	382.19	11.22	11.05	382.56	10.77	10.80
5	20	313.77	38.33	382.56	10.50	10.55	377.20	9.84	9.88
6	25	217.15	47.29	210.31	5.48	5.53	206.35	5.64	5.68
7	22	210.95	40.52	275.32	5.14	5.20	274.60	5.14	5.21
8	24	322.44	49.57	253.77	5.15	5.21	253.15	4.99	5.05
9	24	245.99	40.20	191.28	5.06	5.11	192.76	5.01	5.06
10	19	200.47	32.64	226.69	5.72	5.77	226.54	5.62	5.66
11	22	150.14	35.90	198.05	5.08	5.13	196.79	5.11	5.15
12	23	306.45	42.77	285.91	6.26	6.31	262.07	6.08	6.13
13	24	187.18	32.57	188.74	5.11	5.16	187.87	5.04	5.08
14	19	333.44	35.09	357.26	8.13	8.18	355.03	8.19	8.24
15	24	205.17	52.85	256.76	4.92	4.99	254.54	4.83	4.89

[a] $\overline{Y}_M = \overline{Y}_{M1} = \overline{Y}_{M2}$.

stationary (i.e., spatial correlation does not change within the scene) and isotropy (i.e., spatial correlation is the same in all directions). Although these assumptions are likely only approximately satisfied for areas as large as TM scenes, nevertheless single fitted variograms were used for the entire scene.

4. Results and discussion

The results of the graphical, pixel-level bias analyses were that, although there was heterogeneity of variance among the residuals, there was no indication of bias for any of the four response variables. Although homogeneity of variance would simplify analyses, the variance estimators accommodate heterogeneity of variance.

The effects of the k-value on estimates for the 10-km radius AOIs were minimal. For $1 \leq k \leq 10$, none of the four τ-means (16) or the ψ-mean (17) was more than 5% greater than the minimum value. Therefore, $k=5$ was selected on the basis that it is an approximate minimum sample size for obtaining reliable estimates of σ_i^2.

For $k=5$, variogram models were fit to the standardized residuals for each combination of response variable and permitted number of nearest neighbors associated with subplots of the same FIA plot. There was not sufficient information from the data available to conclude that $\hat{\alpha}_0=0$ from (7) for any of the response variables. Therefore, the variogram model (7a) was simplified to,

$$\hat{\gamma}(h) = \hat{\alpha}_1[1 - \exp(\hat{\alpha}_2 d_{ij})],$$

where $\hat{\alpha}_1$ is the estimate of the variogram sill. The effective range, r, of spatial correlation, defined as the distance at which 95% of the variogram sill is reached, is estimated as,

$$r = \frac{\ln(0.05)}{\hat{\alpha}_2}.$$

Effective ranges of spatial correlation for residuals were relatively small, less than 100 m for all combinations except for PFA when only one neighbor associated with the same FIA plot was permitted; for the latter combination, the effective range of spatial correlation was approximately 155 m.

Estimates obtained when only one neighbor associated with subplots of the same FIA plot was permitted were similar to those obtained when multiple neighbors were permitted (Table 1). For $k=5$, one neighbor associated with subplots of the same FIA plot, and each response variable, $\overline{Y}_P$ was graphed against $\overline{Y}_M$ for each of the 15 AOIs. Generally, the estimates, $\overline{Y}_P$, were along the 1:1 line having intercept 0 and slope 1 (Fig. 2a–d). In addition, with only a few exceptions, vertical lines centered at $\overline{Y}_P$ and extending two probability-based standard errors in either direction intersected the 1:1 line. Root mean square and mean absolute differences were nearly indistinguishable with respect to whether one neighbor or multiple neighbors were permitted for subplots of the same FIA plot (Table 2).

Calculating the variance estimates was computationally intensive because covariances among all pairs of pixel predictions were theoretically necessary. Each 10-km radius circular AOI included approximately 350,000 30-m × 30-m Landsat pixels which meant that approximately 1.2×10^{11} covariance calculations were necessary. To mitigate this computational problem, two steps were taken. First, for each response variable, each AOI was tessellated into squares with side dimensions equal to the range of spatial correlation. Thus, for a given pixel, other pixels for which spatial correlation may be non-zero must be in the square containing the given pixel or in adjacent squares. Therefore, for a given pixel, calculation of covariances among pixel predictions was restricted to pixels in the same square containing the given pixel and adjacent squares. Second, the order in which pixels were selected for calculation of covariances with other pixels was randomized. For each response variable for three AOIs, the accumulated variance estimates, scaled to reflect the total pixels in the AOI, were graphed against the proportion of pixels for which covariances had been calculated. The variance estimates stabilized to within 0.5% of the final estimate by the time covariances for 15% of pixels had been calculated (Fig. 3). For the remaining 12 AOIs, variance estimates were based on a 20% random sample of all pixels in the AOI. These two steps reduced computation time by a factor of more than 50.

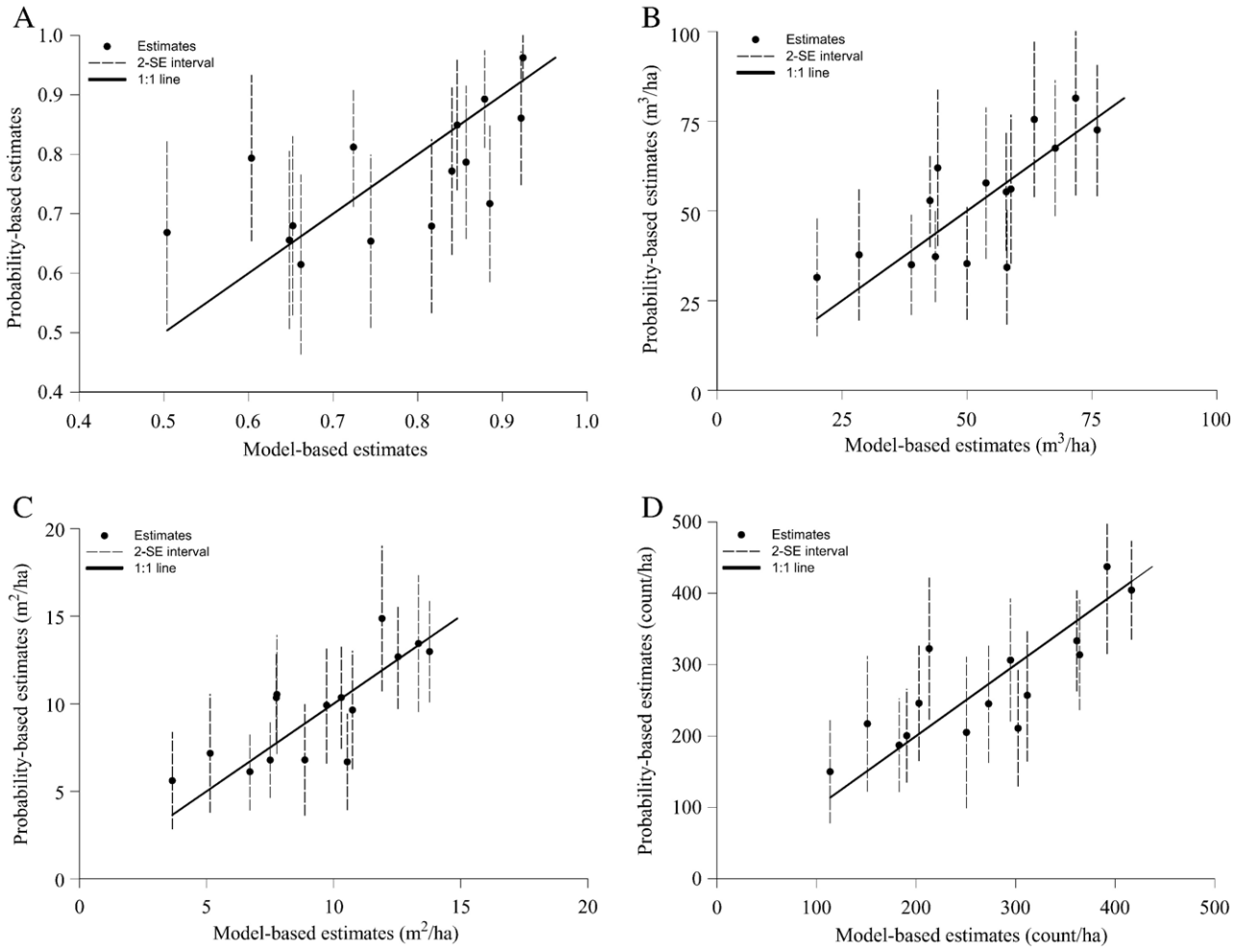

Fig. 2. (A) Comparison of probability-based and model-based (k-NN, $k=5$) estimates for proportion forest area. (B) Comparison of probability-based and model-based (k-NN, $k=5$) estimates for volume (m^3/ha). (C) Comparison of probability-based and model-based (k-NN, $k=5$) estimates for basal area (m^2/ha). (D) Comparison of probability-based and model-based (k-NN, $k=5$) estimates for stem density (tree count/ha).

For each AOI and response variable combination, the standard error estimates were compared for the four combinations resulting from selecting either one or multiple neighbors per plot and considering the k-NN pixel prediction as either an estimate of a mean or as an estimate of a realization from the superpopulation. Although the four estimates were very similar, the standard error estimates corresponding to the k-NN pixel predictions considered as estimates of means were smaller than when predictions were considered as predictions of realizations from the superpopulation. Although this result was as required as per (15c), the differences were small, a result attributed to the relatively small effective ranges of spatial correlation. Also, standard errors were

Table 2
Comparison of estimates for probability-based and k-NN approaches and for one or multiple neighbors per plot

Variable	Number of neighbors permitted per plot			
	1		Multiple	
	Mean squared difference	Mean absolute difference	Mean squared difference	Mean absolute difference
Proportion forest area	0.098	0.078	0.099	0.080
Volume (m^3/ha)	10.82	8.82	11.12	8.95
Basal area (m^2/ha)	1.87	1.45	1.91	1.46
Stem density (count/ha)	51.30	42.38	51.84	42.38

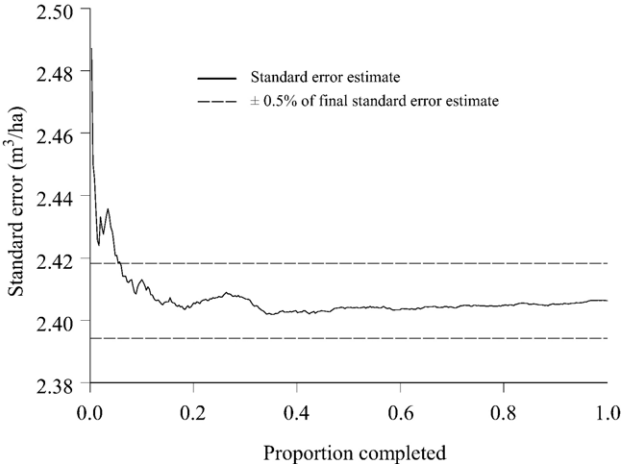

Fig. 3. Scaled standard error estimate versus proportion of pixels for volume, the fifth AOI, multiple neighbors per plot, and k-NN prediction considered as an estimate of the mean.

generally smaller when multiple neighbors associated with the same FIA plot were permitted than when only one neighbor was permitted, although the differences were again slight. This result should not be construed to suggest that multiple neighbors associated with subplots of the same FIA plot were only infrequently selected. For $k=2$, 20% of pixels exhibited this phenomenon; for $k=5$, it was 64%; and for $k=10$, it was 92%. The conclusion is that the reduction in the numerator of the estimate of σ_i^2 from (6a) as a result of this phenomenon was nearly offset by a comparable reduction in the denominator. Overall, coefficients of variation were small; less than 0.011 for PFA, less than 0.036 for VOL, less than 0.052 for BA, and less than 0.030 for D. Variance estimates for PFA were of the same order of magnitude as those reported by McRoberts (2006) for the same AOIs and datasets using a logistic regression model approach.

5. Conclusions

Multiple conclusions may be drawn from this study. First, the computational intensity necessary to calculate the covariances may be mitigated using the techniques described in the first paragraph of the Results and discussion section. Nevertheless, in the absence of faster processing capabilities and more efficient algorithms, the *k*-NN model-based approach may still be too computationally intensive for large areas. Second, the similarity between the estimates obtained using the probability-based and model-based approaches suggests unbiasedness for the *k*-NN areal estimates. Third, spatial correlation had little effect on the variances of estimates, although other applications with greater ranges of spatial correlation may produce different results. Fourth, the effects of permitting only one versus multiple neighbors among the pixels associated with subplots of the same FIA plot were minimal. Fifth, the advantages of the *k*-NN model-based approach include small area estimation, multivariate estimation, and compatible maps as by-products. The variances obtained using the probability-based and model-based approaches cannot be compared in a conceptual sense because of the differences in underlying assumptions and the manner in which variability is generated. Nevertheless, from a practical perspective, disadvantages accruing from the *k*-NN model-based approach do not include loss of precision relative to the probability-based approach.

Finally, the results of the study lead to a 3-part recommendation for obtaining areal estimates of forest attributes using the *k*-NN model-based approach when ranges of spatial correlation are small and the reference set is relatively large. First, consider the *k*-NN prediction to be an estimate of the mean of the distribution of possible observations for the same values of the covariates and use the estimator for $\mathrm{Var}(\bar{Y}_{M1})$; second, permit only one nearest neighbor within the range of spatial correlation of other neighbors; and third, use algorithms similar to those described to reduce computational intensity. The effects of the first two parts of the recommendation further reduce the computational intensity because spatial correlation and covariances may be ignored when estimating σ_i^2, although they must still be considered when estimating $\mathrm{Cov}(\hat{\mu}_i,\hat{\mu}_j)$.

The benefits of the results of this investigation are threefold. First, the utility of the *k*-NN technique has been extended from producing estimates of only means and totals of forest attributes over spatial areas to producing estimates of the uncertainty of estimates of those means and totals. Thus, the statistical significance of differences in estimates of forest attributes for different areas or under different conditions may be assessed. Second, the *k*-NN model-based approach contributes to alleviating the problem of inadequate sample sizes for small area inventory estimation. Third, the *k*-NN model-based approach facilitates simultaneous and consistent mapping and estimation of multiple forest attributes.

This study represents one of the first attempts to derive variance estimators for nearest neighbor methods. As such, there are numerous future research issues. First, investigations should be conducted on the robustness of the superpopulation and variance estimators to violations of assumptions such as unbiasedness, symmetry of the distribution of neighbors in covariate space, and adequacy of the range of reference set observations in covariate space. In addition, the effect of the selection of *k* on variances has not been investigated, nor has the effect of weighting individual neighbors when calculating *k*-NN predictions using (3) rather than the simplified form of (5).

Appendix A *A.1. An estimator for* σ_i^2

The derivation of the estimator for σ_i^2 begins with,

$$s_i^2 = \frac{1}{k}\sum_{j=1}^{k}\left({y_j}^i - \tilde{y}_i\right)^2, \tag{A1}$$

where $\{{y_j}^i: j=1,\ldots,k\}$ is the set of k nearest neighbors to the ith pixel in covariate space and $\tilde{y}_i = \frac{1}{k}\sum_{j=1}^{k}{y_j}^i$ is considered an estimator of μ_i. The statistical expectation of s_i^2 (A1) may be expressed as,

$$\begin{aligned}
E(s_i^2) &= E\left[\frac{1}{k}\sum_{j=1}^{k}\left({y_j}^i-\hat{\mu}_i\right)^2\right] = \frac{1}{k}E\left[\sum_{j=1}^{k}\left({y_j}^i\right)^2 - k\tilde{y}_i^2\right]\\
&= \frac{1}{k}E\left[\sum_{j=1}^{k}\left({y_j}^i\right)^2 - k\left(\frac{1}{k}\sum_{j=1}^{k}{y_j}^i\right)^2\right]\\
&= \frac{1}{k}\left\{E\left[\sum_{j=1}^{k}\left({y_j}^i\right)^2\right] - kE\left(\frac{1}{k^2}\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}{y_{j_1}}^i{y_{j_2}}^i\right)\right\}\\
&= \frac{1}{k}\left\{E\left[\sum_{j=1}^{k}\left(\mu_j+\varepsilon_j\right)^2\right]\right.\\
&\quad\left. - \frac{1}{k}E\left[\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\left(\mu_{j_1}+\varepsilon_{j_1}\right)\left(\mu_{j_2}+\varepsilon_{j_2}\right)\right]\right\}\\
&= \frac{1}{k}\left\{E\left[\sum_{j=1}^{k}\left(\mu_j^2+2\mu_j\varepsilon_j+\varepsilon_j^2\right)\right]\right.\\
&\quad\left. - \frac{1}{k}E\left[\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\left(\mu_{j_1}\mu_{j_2}+\mu_{j_1}\varepsilon_{j_2}+\mu_{j_2}\varepsilon_{j_1}+\varepsilon_{j_1}\varepsilon_{j_2}\right)\right]\right\}.
\end{aligned} \tag{A2}$$

Assuming independence between the μ's and the ε's, (A2) simplifies to,

$$E(s_i^2) = \frac{1}{k}\left[\left(\sum_{j=1}^{k}\mu_j^2\right) + E\left(\sum_{j=1}^{k}\varepsilon_j^2\right) - \frac{1}{k}\left(\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\mu_{j_1}\mu_{j_2}\right) - \frac{1}{k}E\left(\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\varepsilon_{j_1}\varepsilon_{j_2}\right)\right]$$
$$= \frac{1}{k}\left[\left(\sum_{j=1}^{k}\mu_j^2\right) + \left(\sum_{j=1}^{k}\sigma_j^2\right) - \frac{1}{k}\left(\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\mu_{j_1}\mu_{j_2}\right) - \frac{1}{k}\left(\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\sigma_{j_1}\sigma_{j_2}\rho_{j_1 j_2}\right)\right]. \quad \text{(A3)}$$

In the neighborhood of the ith pixel defined by its k nearest neighbors in covariate space, the following approximations are reasonable:

$$\sum_{j=1}^{k}\mu_j^2 \approx k\mu_i^2,\ \sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\mu_{j_1}\mu_{j_2} \approx k^2\mu_i^2,$$

$$\sum_{j=1}^{k}\sigma_j^2 \approx k\sigma_i^2,\ \text{and}\ \sigma_{j_1}\sigma_{j_2} \approx \sigma_i^2.$$

Substituting the approximations into (A3),

$$E(s_i^2) \approx \frac{1}{k}\left[k\mu_i^2 + k\sigma_i^2 - \frac{1}{k}(k^2\mu_i^2) - \frac{1}{k}\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\sigma_i^2\rho_{j_1 j_2}\right]$$
$$= \frac{1}{k}\left(k\sigma_i^2 - \frac{1}{k}\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\sigma_i^2\rho_{j_1 j_2}\right)$$
$$= \sigma_i^2\left(1 - \frac{1}{k^2}\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\rho_{j_1 j_2}\right). \quad \text{(A4)}$$

Thus, an estimator of σ_i^2 is

$$\hat{\sigma}_i^2 \approx \frac{s_i^2}{1 - \frac{1}{k^2}\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\rho_{j_1 j_2}} = \frac{\sum_{j=1}^{k}(y_j{}^i - \tilde{y}_i)^2}{k\left(1 - \frac{1}{k^2}\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\rho_{j_1 j_2}\right)}$$
$$= \frac{\sum_{j=1}^{k}(y_j{}^i - \tilde{y}_i)^2}{k - \frac{1}{k}\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\rho_{j_1 j_2}}. \quad \text{(A5)}$$

A.2. Cov($\hat{\mu}_i$, $\hat{\mu}_j$)

An approximation of the covariance of estimates of arbitrary superpopulation means, $\hat{\mu}_i$ and $\hat{\mu}_j$, may be derived as,

$$\begin{aligned}\mathrm{Cov}(\hat{\mu}_i,\hat{\mu}_j) &= E\left[(\hat{\mu}_i - \mu_i)(\hat{\mu}_j - \mu_j)\right] \\ &= E\left[\left(\frac{1}{k}\sum_{l_i=1}^{k} y_{l_i}{}^i - \mu_i\right)\left(\frac{1}{k}\sum_{l_j=1}^{k} y_{l_j}{}^j - \mu_j\right)\right] \\ &= \frac{1}{k^2}E\left\{\left[\sum_{l_i=1}^{k}(\mu_{l_i} + \varepsilon_{l_i}) - k\mu_i\right] \times \left[\sum_{l_j=1}^{k}(\mu_{l_j} + \varepsilon_{l_j}) - k\mu_j\right]\right\} \\ &= \frac{1}{k^2}E\left\{\left[\left(\sum_{l_i=1}^{k}\mu_{l_i} - k\mu_i\right) + \left(\sum_{l_i=1}^{k}\varepsilon_{l_i}\right)\right] \times \left[\left(\sum_{l_j=1}^{k}\mu_{l_j} - k\mu_j\right) + \left(\sum_{l_j=1}^{k}\varepsilon_{l_j}\right)\right]\right\}\end{aligned} \quad \text{(B1)}$$

where $\{y_{l_i}{}^i: l_i=1, \ldots,k\}$ and $\{y_{l_{ji}}{}^j: l_j=1, \ldots,k\}$ are the sets of k neighbors in the reference set nearest to the ith and jth pixels, respectively, in covariate space. Assuming independence between the μ's and ε's, (B1) may be expressed as,

$$\mathrm{Cov}(\hat{\mu}_i,\hat{\mu}_j) = \frac{1}{k^2}E\left[\left(\sum_{l_i=1}^{k}\mu_{l_i} - k\mu_i\right)\left(\sum_{l_j=1}^{k}\mu_{lj} - k\mu_j\right) + \left(\sum_{l_i=1}^{k}\varepsilon_{l_i}\right)\left(\sum_{l_j=1}^{k}\varepsilon_{l_j}\right)\right].$$

Assuming that in the neighborhoods of the ith and jth pixels $\sum_{l_i=1}^{k}\mu_{l_i} \approx k\mu_i$ and $\sum_{l_{ji}=l}^{k}\mu_{l_j} \approx k\mu_j$, then

$$\mathrm{Cov}(\hat{\mu}_i,\hat{\mu}_j) \approx \frac{1}{k^2}E\left(\sum_{l_i=1}^{k}\sum_{l_j=1}^{k}\varepsilon_{l_i}\varepsilon_{l_j}\right).$$

Further, assuming that in the neighborhoods of the ith and jth pixels $\sigma_{l_i} \approx \sigma_i$ and $\sigma_{l_j} \approx \sigma_j$, then,

$$\mathrm{Cov}(\hat{\mu}_i,\hat{\mu}_j) \approx \frac{\sigma_i\sigma_j}{k^2}\sum_{l_i=1}^{k}\sum_{l_j=1}^{k}\rho_{l_i l_j}. \quad \text{(B2)}$$

A.3. Cov($\hat{y}_i$, $\hat{y}_j$)

An approximation of the covariance of predictions of realizations, $\hat{y}_i$ and $\hat{y}_j$, from arbitrary ith and jth distributions may be derived as,

$$\begin{aligned}\mathrm{Cov}(\hat{y}_i,\hat{y}_j) &= E[(\hat{y}_i - y_i)(\hat{y}_j - y_j)] \\ &= E\left[\left(\frac{1}{k}\sum_{l_i=1}^{k} y_{l_i}{}^{i} - y_i\right)\left(\frac{1}{k}\sum_{l_j=1}^{k} y_{l_j}{}^{j} - y_j\right)\right] \\ &= \frac{1}{k^2}E\left\{\left[\sum_{l_i=1}^{k}(\mu_{l_i}+\varepsilon_{l_i}) - k(\mu_i+\varepsilon_i)\right]\right. \\ &\quad \left.\times\left[\sum_{l_j=1}^{k}(\mu_{l_j}+\varepsilon_{l_j}) - k(\mu_j+\varepsilon_j)\right]\right\} \\ &= \frac{1}{k^2}E\left\{\left[\left(\sum_{l_i=1}^{k}\mu_{l_i} - k\mu_i\right) + \left(\sum_{l_i=1}^{k}\varepsilon_{l_i} - k\varepsilon_i\right)\right]\right. \\ &\quad \left.\times\left[\left(\sum_{l_j=1}^{k}\mu_{l_j} - k\mu_j\right) + \left(\sum_{l_j=1}^{k}\varepsilon_{l_j} - k\varepsilon_j\right)\right]\right\}\end{aligned} \tag{C1}$$

where $\{y_{l_i}{}^{i}: l_i = 1, \ldots, k\}$ and $\{y_{l_j}{}^{j}: l_j = 1, \ldots, k\}$ are the sets of k neighbors in the reference set nearest to the ith and jth pixels, respectively, in covariate space. Assuming independence between the μ's and ε's, (C1) may be expressed as,

$$\begin{aligned}\mathrm{Cov}\left(\hat{y}_i,\hat{y}_j\right) = \frac{1}{k^2}E\left[\left(\sum_{l_i=1}^{k}\mu_{l_i} - k\mu_i\right)\left(\sum_{l_i=1}^{k}\mu_{l_j} - k\mu_j\right)\right. \\ \left.+\left(\sum_{l_i=1}^{k}\varepsilon_{l_i} - k\varepsilon_i\right)\left(\sum_{l_j=1}^{k}\varepsilon_{l_j} - k\varepsilon_j\right)\right].\end{aligned}$$

Assuming that in the neighborhoods of the ith and jth pixels $\sum_{l_i=1}^{k}\mu_{l_i} \approx k\mu_i$ and $\sum_{l_j=1}^{k}\mu_{l_j} \approx k\mu_j$, then

$$\begin{aligned}\mathrm{Cov}\left(\hat{y}_i,\hat{y}_j\right) &\approx \frac{1}{k^2}E\left[\left(\sum_{l_i=1}^{k}\varepsilon_{l_i} - k\varepsilon_i\right)\left(\sum_{l_j=1}^{k}\varepsilon_{l_j} - k\varepsilon_j\right)\right] \\ &= \frac{1}{k^2}E\left(\sum_{l_i=1}^{k}\sum_{l_j=1}^{k}\varepsilon_{l_i}\varepsilon_{l_j} - k\sum_{l_i=1}^{k}\varepsilon_{l_i}\varepsilon_j - k\sum_{l_j=1}^{k}\varepsilon_i\varepsilon_{l_j} + k^2\varepsilon_i\varepsilon_j\right).\end{aligned}$$

Further, assuming that in the neighborhoods of the ith and jth pixels $\sigma_{l_i} \approx \sigma_i$ and $\sigma_{l_j} \approx \sigma_j$, then,

$$\mathrm{Cov}\left(\hat{y}_i,\hat{y}_j\right) \approx \frac{\sigma_i\sigma_j}{k^2}\left(\sum_{l_i=1}^{k}\sum_{l_j=1}^{k}\rho_{l_il_j} - k\sum_{l_i=1}^{k}\rho_{l_ij} - k\sum_{l_j=1}^{k}\rho_{il_j} + k^2\right). \tag{C2}$$

References

Ardö, J. (1992). Volume quantification of coniferous forest compartments using spectral radiance recorded by Landsat Thematic Mapper. *International Journal of Remote Sensing*, *13*, 1779–1786.

Aubry, P., & Debouzie, D. (2000). Geostatistical estimation variance for the spatial mean in two-dimensional systematic sampling. *Ecology*, *81*, 543–553.

Bechtold, W. A., & Patterson, P. L. (Eds.). (2005). *The enhanced Forest Inventory and Analysis program — national sampling design and estimation procedures. General technical report SRS-80*. Asheville, North Carolina: Southern Research Station, USDA Forest Service 85 pp.

Cochran, W. G. (1977). *Sampling techniques*, 3rd ed. New York: Wiley.

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.

Draper, N., & Smith, H. (1981). *Applied regression analysis*, 2nd edition. New York: Wiley.

Dungan, J. L. (1998). Spatial prediction of vegetation quantities using ground and image data. *International Journal of Remote Sensing*, *19*, 267–285.

Fazakas, Z., & Nilsson, M. (1996). Volume of forest cover estimation over southern Sweden using AVHRR data calibrated with TM data. *International Journal of Remote Sensing*, *17*, 1701–1709.

Franco-Lopez, H., Ek, A. R., & Bauer, M. E. (2001). Estimation and mapping of forest stand density, volume, and cover type using the *k*-nearest neighbors method. *Remote Sensing of Environment*, *77*, 251–1709.

Graybill, F. A. (1961). *An introduction to linear statistical models.* New York: McGraw-Hill.

Gregoire, T. G. (1998). Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, *28*, 1429–1447.

Gumpertz, M. L, Wu, C., & Pye, J. M. (2000). Logistic regression for southern pine beetle outbreaks with spatial and temporal autocorrelation. *Forest Science*, *46*, 95–107.

Halme, M., & Tomppo, E. (2001). Improving the accuracy of multisource forest inventory estimates by reducing plot location error — a multicriteria approach. *Remote Sensing of Environment*, *78*, 321–327.

Hansen, M. H., & Wendt, D. G. (2000). Using classified Landsat Thematic Mapper data for stratification in a statewide forest inventory. In R. E. McRoberts, G. A. Reams, & P. C. Van Deusen (Eds.), *Proceedings of the first annual forest inventory and analysis symposium. General technical report, NC-213*. (pp.20–27) St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station.

Hansen, M. H., Madow, W. G., & Tepping, B. J. (1978). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association*, *78*, 776–793.

Hassani, B. T., LeMay, V., Marshall, P. L., Temsegen, H., & Zumrawi, A. -A. (2004). Regeneration imputation models for complex stands of southeaster British Columbia. *Forestry Chronicle*, *80*, 271–278.

Holmström, H., & Fransson, J. E. S. (2003). Combining remotely sensed optical and radar data in *k*-NN estimation of forest variables. *Forest Science*, *4*, 409–418.

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*, 829–840.

Katila, M., & Tomppo, E. (2001). Selecting estimation parameters for the Finnish multisource National Forest Inventory. *Remote Sensing of Environment*, *76*, 16–32.

Katila, M., & Tomppo, E. (2002). Stratification by ancillary data in multisource forest inventories employing *k*-nearest neighbour estimation. *Canadian Journal of Forest Research*, *3*, 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. *Proceedings of the symposium on machine processing of remotely sensed data* (pp. 41–51). West Lafayette, IN: Purdue University.

Kendall, M. G., & Buckland, W. R. (1982). *A dictionary of statistical terms*, 4th ed. London: Longman Group.

Kim, H. -J., & Tomppo, E. (2006). Model-based prediction error uncertainty estimation for *k*-NN method. *Remote Sensing of Environment*, *104*, 257–263.

Lachenbruch, P. A., & Mickey, M. R. (1986). Estimation of error rates in discriminant analysis. *Technometrics*, *10*, 1–11.

Liknes, G. C., Nelson, M. D., & McRoberts, R. E. (2004). Evaluating classified MODIS satellite imagery as a stratification tool. In H. T. Mowrer, R. E. McRoberts, & P. C. Van Deusen (Eds.), *Joint proceedings of the sixth*

international symposium on spatial accuracy in natural resources and environmental sciences and the fifteenth annual conference of the The International Environmetrics Society.

Mäkelä, H., & Pekkarinen, A. (2001). Estimation of timber volume at the sample plot level by means of image segmentation and Landsat TM imagery. *Remote Sensing of Environment*, *77*, 66−77.

Mäkelä, H., & Pekkarinen, A. (2004). Estimation of forest stand volume by Landsat TM imagery and stand-level field-inventory data. *Forest Ecosystem and Management*, *196*, 245−255.

McRoberts, R. E. (2006). A model-based approach to estimating forest area. *Remote Sensing of Environment*, *103*, 56−66.

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., Holden, G. R., Nelson, M. D., Liknes, G. C., & Gormanson, D. D. (2006). Using satellite imagery as ancillary data for increasing the precision of estimates for the Forest Inventory and Analysis program of the USDA Forest Service. *Canadian Journal of Forest Research*, *36*, 2968−2980.

McRoberts, R. E., Nelson, M. D., & Wendt, D. G. (2002a). 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., Wendt, D. G., Nelson, M. D., & Hansen, M. H. (2002b). Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. *Remote Sensing of Environment*, *81*, 36−44.

Moeur, M., & Stage, A. R. (1995). Most Similar Neighbor: an improved sampling inference procedure for natural resource planning. *Forest Science*, *41*, 337−359.

Nilsson, M., Holm, S., Reese, H., Wallerman, J., & Engberg, J. (2005). Improved forest statistics from the Swedish National Forest Inventory by combining field data and optical satellite data using post-stratification. *Proceedings of ForestSat 2005, 31 May–03 June 2005, Borås, SwedenRapport*, *vol. 8a*. (pp.22−26) Skogsstyrelsen: National Board of Forestry.

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.

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the great plains with ERTS. *Proceedings of the Third ERTS Symposium, NASA SP-351*, *vol. 1*. (pp.309−317) Washington, DC: NASA.

Tokola, T., Pitkänen, J., Partinen, S., & Muinonen, E. (1996). Point accuracy of non-parametric method in estimation of forest characteristics with different satellite materials. *International Journal of Remote Sensing*, *17*, 2333−2351.

Tomppo, E. (1988). Standwise forest variate estimation by means of satellite images. *Proceedings of the IUFRO S4.02.05 Meeting, Aug. 29–Sept. 2, 1988 Forest Station Hyytiälä, FinlandResearch Notes*, *vol. 21*. (pp.103−111) : University of Helsinki, Department of Forest Mensuration and Management, 951-45-4820-5.

Tomppo, E. (1991). Satellite image-based national forest inventory of Finland. *International Archives of Photogrammetry and Remote Sensing*, *28*, 419−424.

Tomppo, E., Goulding, C., & Katila, M. (1999). Adapting Finnish multisource inventory techniques to the New Zealand pre-inventory harvest. *Scandinavian Journal of Forest Research*, *14*, 182−192.

Tomppo, E., & Halme, M. (2004). Using course scale forest variables as ancillary information and weighting of variables in *k*-NN estimation: a genetic algorithm approach. *Remote Sensing of Environment*, *92*, 1−20.

Tomppo, E., Nilsson, M., Rosengren, M., Aalto, P., & Kennedy, P. (2002). Simultaneous use of Landsat-TM and IRS-1C WiFS data in estimating large area tree stem volume and aboveground biomass. *Remote Sensing of Environment*, *82*, 156−171.

Trotter, C. M., Drymond, J. R., & Goulding, C. J. (1997). Estimation of timber volume in a coniferous plantation forest using Landsat TM. *International Journal of Remote Sensing*, *18*, 2209−2223.

Yang, L., Homer, C. G., Hegge, K., Huang, C., Wylie, B., & Reed, B. (2001). A Landsat 7 scene selection strategy for a National Land Cover Database. *Proceedings of the IEEE 2001 International Geoscience and Remote Sensing Symposium, Sydney, Australia* CD-ROM, 1 disc.