Remote Sensing of Environment 114 (2010) 1017–1025

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Probability- and model-based approaches to inference for proportion forest using satellite imagery as ancillary data

Ronald E. McRoberts

Northern Research Station, U.S. Forest Service, Saint Paul, Minnesota, USA

ARTICLE INFO

Article history:
Received 2 April 2009
Received in revised form 18 December 2009
Accepted 20 December 2009

Keywords:
Forest inventory
Small area estimation
Inference
Stratification

ABSTRACT

Estimates of forest area are among the most common and useful information provided by national forest inventories. The estimates are used for local and national purposes and for reporting to international agreements such as the Montréal Process, the Ministerial Conference on the Protection of Forests in Europe, and the Kyoto Protocol. The estimates are usually based on sample plot data and are calculated using probability-based estimators. These estimators are familiar, generally unbiased, and entail only limited computational complexity, but they do not produce the maps that users are increasingly requesting, and they generally do not produce sufficiently precise estimates for small areas. Model-based estimators overcome these disadvantages, but they may be biased and estimation of variances may be computationally intensive. The study objective was to compare probability- and model-based estimators of mean proportion forest using maps based on a logistic regression model, forest inventory data, and Landsat imagery. For model-based estimators, methods for evaluating bias and reducing the computational intensity were also investigated. Four conclusions were drawn: the logistic regression model exhibited no serious lack of fit to the data; all the estimators produced comparable estimates for mean proportion forest, except for small areas; probability-based inferences enhanced using maps produced increased precision; and the computational intensity associated with estimating variances for model-based estimators can be greatly reduced with no detrimental effects.

Published by Elsevier Inc.

1. Introduction

Forest area is one of the mostly commonly assessed variables by national forest inventories (NFI). It is also an indicator specified by the two primary forest sustainability agreements, the Montréal Process (2005) and the Ministerial Convention on the Protection of Forests in Europe (2009). In addition, parties to the United Nations Framework Convention on Climate Change (UNFCCC, 2007) and the Kyoto Protocol (1997) submit annual reports of emissions for six land use categories of which one is forest. Of particular importance, forest is the only category whose resources can be managed to produce a positive effect on the greenhouse gas balance (Cienciala et al., 2008).

NFI sample plot data are the primary source of information for estimating forest area. Traditionally, NFIs have reported estimates calculated using probability-based estimators, also characterized as design-based estimators, but generally only for large areas such as countries, regions within countries, states, and provinces. With probability-based approaches estimates are based on sample plot observations, and the validity of inference is based on the probabilistic nature of the sampling design. The primary advantages of probability-based estimators are that they are familiar and are generally unbiased. The disadvantages are that they do not produce the maps that users increasingly request; they do not necessarily produce estimates consistent with maps; and, because of small sample sizes, they do not yield sufficiently precise estimates for small areas.

Model-based approaches, sometimes characterized as model-dependent approaches (Hansen et al., 1983), use predictions based on models and ancillary variables to produce estimates. The primary advantages of model-based estimators are that they produce maps as by-products, estimates that are consistent with the maps in the sense that they represent the aggregation of population unit predictions, and more precise estimates for small areas. Their primary disadvantages are that they are not necessarily unbiased, and they are often computationally intensive.

Stratified and model-assisted estimators can use maps based on models and ancillary data to increase precision. However, because the validity of the inferences obtained using these estimators is still based on probability samples, they are characterized as probability-based. In the sense that inferences based on these estimators use models but depend on probability samples for validity, they may be considered to represent a middle-ground between sampling random sampling and model-based estimators.

The objective of the study was to compare approaches to inference using probability- and model-based estimators of mean proportion

E-mail address: rmcroberts@fs.fed.us.

0034-4257/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.rse.2009.12.013

forest. The emphasis was on mean proportion forest rather than forest area, because the latter can be estimated as the product of mean proportion forest and total area which is usually known from other sources. A logistic regression model was used with forest inventory sample plot data and Landsat Thematic Mapper (TM) imagery to construct a map of estimates of the probability of forest for each image pixel. Maps based on the estimates were used with stratified and model-assisted estimators to enhance probability-based inference and served as the basis for model-based inference. Issues of bias assessment and computational intensity were addressed for the model-based approach. Comparisons of inferential approaches were based on comparisons of estimates of mean proportion forest and standard errors of the estimates for areas of interest (AOI) of varying sizes.

2. Data

The study area was defined by the portion of the row 27, path 27, Landsat scene in northern Minnesota, USA (Fig. 1). Land use for the study area consists of forest land dominated by aspen-birch and spruce-fir associations, agriculture, wetlands, and water. TM imagery was acquired for three dates corresponding to early, peak, and late seasonal vegetative stages (Yang et al., 2001), April 2000, July 2001, and November 1999. Preliminary analyses indicated that the normalized difference vegetation index (NDVI) (Rouse et al., 1973) and the tasseled cap (TC) transformations (brightness, greenness, and wetness) (Crist & Cicone, 1984; Kauth & Thomas, 1976) were superior to both the spectral band data and principal component transformations with respect to predicting the probability of forest cover. Thus, 12 TM-based variables were used, NDVI and the three TC transformations for each of the three image dates. Centers for 14 circular AOIs were selected to provide a systematic sample of the study area.

Fig. 1. Study area.

Forest inventory data for 2232 permanent field plots observed between the beginning of 1999 and the end of 2003 were available for the study area. Of these plots, 731 were located within 15 km of the 14 AOI centers. The plot locations had been established by the Forest Inventory and Analysis (FIA) program of the U.S. Forest Service using an equal probability sampling design (Bechtold & Patterson, 2005; McRoberts et al., 2005). Each plot consists of four 7.32-m (24-ft) radius circular subplots, and 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° from the center of the central subplot. In general, locations of forested or previously forested plots are determined using global positioning system receivers, whereas locations of non-forested plots are verified using aerial imagery and digitization methods. For this study, only data for the central subplot of each plot were used to avoid issues related to lack of independence in the selection of locations of subplots of the same plot.

Field crews visually estimate the proportions of subplot areas for all land use conditions. Subplot estimates of proportion forest area are obtained by collapsing land use conditions into forest and non-forest classes consistent with the FIA definition of forest land: area of at least 0.4 ha (1.0 ac), external crown-to-crown width of at least 36.58 m (120 ft), stocking of at least 10%, and forest land use. Of the 2232 central subplots, 568 were completely non-forested, 13 were partially non-forested and partially forested, and 1651 were completely forested. Of the 731 central subplots with centers within 15 km of the 14 AOI centers, 176 were completely non-forested, 13 were partially non-forested and partially forested, and 542 were completely forested. The spatial configuration of the FIA subplots with centers separated by 36.58 m and the 30-m × 30-m spatial resolution of the Landsat imagery permits individual subplots to be associated with individual image pixels.

3. Methods

All analyses were based on three underlying assumptions: (1) a finite population consisting of N units in the form of 30-m × 30-m Landsat pixels, (2) an equal probability sample of n population units in the form of observations of the central subplots of FIA plots, and (3) availability of ancillary data in the form of the 12 Landsat-based spectral transformations for each population unit. Although the subplot area of 167.87 m^2 is approximately 19% of the 900 m^2 pixel area, subplot observations were assumed to characterize entire pixels. In the following sections, the terms population unit and pixels are used interchangeably.

3.1. Logistic regression model

Multiple approaches for estimating proportion forest for image pixels may be used including, but not limited to, maximum likelihood (Magnussen et al., 2001), discriminant analysis (Tomppo et al., 2001), neural networks (Liu et al., 2003), nearest neighbors (McRoberts et al., 2002b; Tomppo et al., 2009), and various forms of regression. For data similar to those used for this study, a logistic regression model was previously found to produce positive results. McRoberts et al. (2006) constructed a map of the probability of forest cover using a logistic regression model with FIA data and TM imagery and used the map to construct strata for use with a stratified estimator. A similar map was constructed using the same methods and served as the basis for a model-based approach to estimating forest area (McRoberts, 2006). Finally, McRoberts (2009) compared nearest neighbors, discriminant analysis, and multinomial logistic regression approaches for constructing categorical forest attribute maps and found the logistic regression approach to be superior.

For binomial logistic regression, the probability, μ_i, of forest cover for the ith pixel was estimated using a logistic regression model of the form,

$$\mu_i = E(y_i) + f(X_i;\beta) = \frac{\exp\left(\sum_{j=1}^{J} \beta_j x_{ij}\right)}{1 + \exp\left(\sum_{j=1}^{J} \beta_j x_{ij}\right)}. \quad (1)$$

where exp(.) is the exponential function, β is vector of parameters to be estimated, x_{ij} is the value of the jth spectral transformation for the ith pixel, and E(.) is statistical expectation. For this study, the model parameters were estimated by maximizing the likelihood,

$$L = \prod_{i=1}^{n} \mu_i^{y_i}(1-\mu_i)^{1-y_i},$$

using data for the 2219 central FIA subplots with centers in the TM scene that were completely non-forested ($y_i = 0$) or completely forested ($y_i = 1$) The variance of $\hat{\mu}_i$ is estimated as,

$$\hat{Var}(\hat{\mu}_i) = \hat{Cov}(\hat{\mu}_i, \hat{\mu}_i) = Z_i' V_{\hat{\beta}} Z, \quad (2)$$

and the covariance between $\hat{\mu}_i$ and $\hat{\mu}_j$ is estimated as,

$$\hat{Cov}(\hat{\mu}_i, \hat{\mu}_j) = Z_i' V_{\hat{\beta}} Z_j, \quad (3)$$

where

$$z_{ij} = \frac{\partial f(X_i;\hat{\beta})}{\partial \beta_j},$$

and $V_{\hat{\beta}}$ is the estimated covariance matrix for the parameter estimates. $V_{\hat{\beta}}$ may be calculated as,

$$V_{\hat{\beta}} = Z' V^{-1} Z,$$

where V is a diagonal matrix with elements, $v_{ii} = \hat{\mu}_i(1-\hat{\mu}_i)$. These variance estimators are approximations based on second-order Taylor series expansions and, depending on the degree of model misspecification, may be biased.

3.2. Probability-based estimators

Properties of probability-based estimators derive from the probabilities of selection of population units into the sample, thus the characterization of these estimators as probability-based (Hansen et al., 1983). Although probability-based inference is also characterized as design-based inference, the term *design* is not well-defined in the sense that it may refer simply to the selection of population units into the sample or additionally to the entire inferential process (Kendall & Buckland, 1982). Further, depending on the application, valid inferences do not necessarily require sampling designs with probabilistic components (Gregoire, 1998).

Probability-based inference is based on three assumptions: (1) population units are selected for the sample using a probability-based randomization scheme; (2) the probability of selection for each population unit is positive and is either known or can be estimated; and (3) the observation of the response variable for each population unit is a constant. Estimators are derived to correspond to sampling designs and typically are unbiased, meaning that the expectation of the estimator, $\hat{\mu}$, over all samples that could be obtained with the sampling design is the population parameter; i.e., $E(\hat{\mu}) = \mu$. However, the estimate obtained with any particular sample may deviate considerably from the true value of the population parameter.

3.2.1. Simple random sampling estimator

For simple random sampling (SRS) designs, the simplest approach to probability-based inference for mean proportion forest, m, is to use the familiar SRS estimators,

$$\hat{\mu}_{SRS} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (4)$$

and

$$\hat{Var}(\hat{\mu}_{SRS}) = \frac{\sum_{i=1}^{n}\left(y_i - \hat{\mu}_{SRS}\right)^2}{n(n-1)}, \quad (5)$$

where i indexes the n sample observations and y_i is the observation for the ith population unit. The primary advantages of the SRS estimators are that they are intuitive, simple, and unbiased when used with an SRS design; the disadvantage is that variances are frequently large, particularly for small sample sizes. Although the variance estimator (Eq. 5) may be biased when used with systematic sampling, it is conservative in the sense that it over-estimates the variance (Särndal et al., 1992). In addition, for this study, finite population correction factors are ignored because of the small sampling intensity, one 168-m^2 plot per 1200 ha of land area.

3.2.2. Stratified estimator

For areal estimation, the stratified estimator can use a map in the form of either categorical or continuous predictions for satellite image pixels to increase the precision of estimates and thereby enhance the inference (McRoberts et al., 2002a,b, 2006). However, because the validity of the inference is still based on probabilities of selection of population units into the sample, and observations are still assumed to be constant for each population unit, inferences obtained using the stratified estimator are characterized as probability-based. The essence of stratified estimation is to assign sampled population units to groups or strata, calculate within stratum sample means and variances, and then calculate a weighted average of the within stratum estimates where the weights are proportional to the stratum sizes. If the stratification is effective, the stratified variance estimate will be less than the SRS variance estimate. Of considerable importance, stratified approaches to estimation may produce reductions in variances even when stratified sampling is not used, such as the case when all sampling units are established in permanent locations.

Stratified estimation requires accomplishment of two tasks: (1) calculation of the stratum weights as the relative proportions of the population area corresponding to strata, and (2) assignment of sample units to a single stratum. The first task is accomplished by calculating the stratum weights as proportions of population units (pixels) in strata. The second task is accomplished for this study by assigning the central FIA subplots to strata on the basis of the stratum assignments of the population units (pixels) containing the subplot centers. Stratified estimates are calculated using methods described by Cochran (1977):

$$\hat{\mu}_{Str} = \sum_{h=1}^{H} w_h \hat{\mu}_h, \quad (6)$$

and

$$\hat{Var}\left(\hat{\mu}_{Str}\right) = \sum_{h=1}^{H} w_h^2 \frac{\hat{\sigma}_h^2}{n_h}, \quad (7)$$

where

$$\hat{\mu}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi},$$

$$\hat{\sigma}_h^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}\left(y_{hi} - \hat{\mu}_h\right)^2,$$

$h=1,\ldots,H$ denotes strata; y_{hi} is the ith observation in the hth stratum; w_h is the weight for the hth stratum; n_h is the number of subplots assigned to the hth stratum; $\hat{\mu}_h$ and $\hat{\sigma}_h^2$ are the sample estimates of the within stratum mean and variance, respectively; and *Str* denotes the stratified estimator. As for the SRS variance estimators, these stratified variance estimators are biased when used with systematic sampling, but again they produce conservative estimates. Also, as for the SRS variance estimators, finite population correction factors are ignored because of the small sampling intensity.

Stratified estimation is effective in increasing precision when the variables on which the stratification is based are closely related to the estimation variable, when sample units can be accurately registered to the stratification, when the dates of the sample observations and the imagery underlying the stratification are similar, and when misclassification errors in the underlying classification are minimal (McRoberts et al., 2006). Although misclassification errors affect the utility of the stratification for increasing precision, they do not contribute to bias in the estimator.

3.2.3. Model-assisted, difference estimator

Model-assisted estimators use models based on ancillary data to enhance inferences but rely on the probability sample for validity. A naïve model-assisted estimator, $\hat{\mu}_{naive}$, of mean proportion forest is the mean over all population units of sample observations where they are available and model predictions where observations are not available:

$$\hat{\mu}_{naive} = \frac{1}{N}\left(\sum_{i=1}^{n} y_i + \sum_{i=n+1}^{N} \hat{y}_i\right) \tag{8}$$

where, without loss of generality, the population units are indexed so that $i=1,2,\ldots,$ n denotes the population units selected for the sample, and $i=n+1, n+2,\ldots,N$ denotes the population units not selected for the sample. In addition, the notation $\hat{y}_i$ is used instead of $\hat{\mu}_i$ because for probability-based approaches the model predicts the constant value for a population unit rather than the mean of observations for the same combination of values for the model independent variables. The expectation of $\hat{\mu}_{naive}$ is,

$$\begin{aligned} E\left(\hat{\mu}_{naive}\right) &= E\left[\frac{1}{N}\left(\sum_{i=1}^{n} y_i + \sum_{i=n+1}^{N} \hat{y}_i\right)\right] \\ &= E\left(\frac{1}{N}\sum_{i=1}^{N} y_i\right) + E\left[\frac{1}{N}\sum_{i=n+1}^{N}\left(\hat{y}_i - y_i\right)\right] \\ &= \mu + Bias\left(\hat{\mu}_{naive}\right), \end{aligned}$$

where the bias may be estimated as,

$$\hat{Bias}(\hat{\mu}_{naive}) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right). \tag{9}$$

Therefore, the model-assisted difference (*Dif*) estimator (Baffetta et al., 2009; Särndal et al., 1992, pp. 221-225) is defined as the difference between the naïve estimator (Eq. 8) and its bias estimator (Eq. 9),

$$\hat{\mu}_{Dif} = \frac{1}{N}\left(\sum_{i=1}^{n} y_i + \sum_{i=n+1}^{N} \hat{y}_i\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right),$$

which can be approximated as,

$$\hat{\mu}_{Dif} = \frac{1}{N}\sum_{i=1}^{N} \hat{y}_i - \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right), \tag{10}$$

(Appendix A) with variance that can be approximated as,

$$\hat{Var}\left(\hat{\mu}_{Dif}\right) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2, \tag{11}$$

(Appendix B).

The primary advantage of model-assisted estimators is that they capitalize on the relationship between the sample observations and their model predictions to reduce the variance of the population parameter estimate. In this regard, they are potentially preferable to the stratified estimator because they use the actual predictions whereas the stratified estimator aggregates the predictions into a small number of classes or strata. However, as with all probability-based estimators, model-assisted estimators suffer the effects of small sample sizes when used for small area estimation.

3.3. Model-based inference

The assumptions underlying model-based inference differ considerably from the assumptions underlying probability-based inference. First, the observation for a population unit is a random variable whose value is considered a realization from a distribution of possible values, rather than a constant as is the case for probability-based inference. The conceptual framework with a distribution of possible values for each unit in a finite population is characterized as a superpopulation, and model-based inference is occasionally characterized as superpopulation inference (Graubard & Korn, 2002). Second, the basis for a model-based inference is the model, not the probabilistic nature of the sample as is the case for probability-based inference. In fact, purposive, non-probability samples may produce entirely valid model-based inferences. For example, the sample may be selected to maximize the precision of the model parameter estimates or the precision of model predictions. However, a probability sample provides modest assurance that the ranges of values of independent variables in the sample data are similar to the ranges in the population to which the model is applied. Randomization for probability-based inference enters through the random selection of population units into the sample, whereas randomization for model-based inference enters through the random realizations from the distributions for population units.

Current approaches to model-based inference originated in the context of survey sampling and can be attributed to Mátern (1960), Brewer (1963), and Royall (1970). Given the origins of model-based inference in survey sampling, it is not surprising that forestry applications have often been in the context of forest inventory (Rennolls, 1982; Gregoire, 1998; Kangas & Maltamo, 2006; Mandallaz, 2008). An important aspect of the model-based approach is that when the model is correct, the estimator is unbiased (Lohr, 1999); however, when the model is misspecified, the adverse effects on inference may be substantial (Hansen et al., 1983; Royall & Herson, 1973). Thus, much of the reported research on model-based inference has focused on selection of sampling designs and estimators that are robust to model misspecification, particularly for linear models (Chambers, 2003; Valliant et al., 2000). Despite a few examples for nonlinear models such as Valliant (1985), considerable work remains when the models are nonlinear (Valliant et al., 2000).

In the context of model-based inference, the mean and standard deviation of the distribution of Y for the ith population unit may be denoted μ_i and σ_i, respectively. Thus, an observation of Y for the ith unit is expressed as,

$$y_i = \mu_i + \varepsilon_i,$$

where ε_i is the random deviation of the observation, y_i, from its mean, μ_i, which is expressed mathematically as,

$$\mu_i = f(X_i;\beta).$$

The model,

$$y_i = f(X_i;\beta) + \varepsilon_i,$$

is fit to the sample data where $\boldsymbol{X}_i$ is the vector of ancillary variables for the ith population unit, $\boldsymbol{\beta}$ is a vector of parameters to be estimated, and ε_i is the random residual.

Two model-based estimators for the population parameter, μ, mean proportion forest, may be considered. First, an estimator may be based on the set of estimates $\{\hat{\mu}_i,\ i=1,\ 2,...,N\}$ of the means for individual population units,

$$\hat{\mu}_{Mod} = \frac{1}{N}\sum_{i=1}^{N}\hat{\mu}_i = \frac{1}{N}\sum_{i=1}^{N} f\left(X_i;\hat{\beta}\right). \tag{12}$$

Second, an estimator may be based on the set of predictions, $\{\hat{y}_i,\ i=1,\ 2,...,N\}$, of random realizations,

$$\bar{\hat{y}}_{Mod} = \frac{1}{N}\sum_{i=1}^{N}\hat{y}_i = \frac{1}{N}\sum_{i=1}^{N} f\left(X_i;\hat{\beta}\right). \tag{13}$$

Both estimators produce the same estimate, because the best prediction of an observation from the distribution for a population unit is the mean of the distribution. However,

$$\hat{Var}(\hat{\mu}_{Mod}) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{Cov}\left(\hat{\mu}_i,\hat{\mu}_j\right), \tag{14}$$

whereas,

$$\begin{aligned}\hat{Var}(\bar{\hat{y}}_{Mod}) &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{Cov}\left(\hat{\mu}_i+\varepsilon_i,\hat{\mu}_j+\varepsilon_j\right) \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{Cov}(\hat{\mu}_i,\hat{\mu}_j) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{Cov}(\varepsilon_i,\varepsilon_j)\end{aligned} \tag{15}$$

under the assumption that the $\hat{\mu}_i$s and the ε_is are independent. In these expressions, $\hat{Var}(\hat{\mu}_i) = \hat{Cov}(\hat{\mu}_i,\ \hat{\mu}_i)$ is the uncertainty of each pixel estimate, $\hat{\mu}_i$; $\hat{Cov}(\hat{\mu}_i,\hat{\mu}_j)$, $i \neq j$, reflects the fact that estimates for different pixels are correlated because they are based on the same sample data; and $\hat{Cov}(\varepsilon_i,\ \varepsilon_j)$ reflects the fact that the εs are correlated, likely in a spatial manner. This spatial covariance may further be expressed as,

$$Cov(\varepsilon_i,\varepsilon_j) = \sigma_i\sigma_j\rho_{ij},$$

where ρ_{ij} is the spatial correlation which is assumed to decrease monotonically with respect to the distance between the ith and jth pixels. Thus, for the binary logistic model, the second component of Eq. (15) may be expressed as,

$$\begin{aligned}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{Cov}(\varepsilon_i,\varepsilon_j) &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\sigma}_i\hat{\sigma}_j\hat{\rho}_{ij} \\ &< \frac{0.25\,m}{N}\end{aligned} \tag{16}$$

where m is the number of population units over which ρ_{ij} is non-negligible, and $\max(\hat{\sigma}_i\hat{\sigma}_j\hat{\rho}_{ij}) \leq \max(\hat{\sigma}_i^2) = \max[\hat{\mu}_i(1-\hat{\mu}_i)] \leq 0.25$. For large N and $\frac{m}{N}$ small, the quantity expressed by Eq. (16) becomes negligible, which means that $\mathrm{Var}(\bar{\hat{y}}_{Mod}) \approx \mathrm{Var}(\hat{\mu}_{Mod})$. Therefore, because the estimators expressed by Eqs. (12) and (13) produce the same estimate, and because the estimates obtained from Eqs. (14) and (15) are the same for large N and small $\frac{m}{N}$, only $\hat{\mu}_{Mod}$ needs to be considered for most applications. Based on variograms using data similar to those used for this study, McRoberts (2006) estimated ranges of spatial correlation to be on the order of 100–120 m which corresponds to $m \approx 50$ and produce $\frac{m}{N} < 0.015$ for an AOI of radius 1 km and $\frac{m}{N} < 0.0006$ for an AOI of radius 5 km. McRoberts (2006) also reported precision estimates based on both Eqs. (14) and (15), found the differences to be negligible, but did not attribute the negligible differences to small $\frac{m}{N}$.

A disadvantage of model-based estimators is that calculation of $\hat{Var}(\hat{\mu}_{Mod})$ using Eq. (14) is computationally intensive because of the double summation and the large number of pixels for even relatively small AOIs. For example, an AOI of radius 15 km includes centers for approximately 7.85×10^5 TM pixels meaning that the number of covariance calculations necessary for Eq. (14) is of the order of 10^{10}. However, Eq. (14) is just a two-dimensional mean over all units in the AOI and can be approximated by sampling from AOIs. Using only units (pixels) located at the intersections of an equally-spaced, two-dimensional, perpendicular grid superimposed on the AOI,

$$\hat{Var}(\hat{\mu}_{Mod}) \approx \frac{1}{n_{grid}^2}\sum_{i=1}^{n_{grid}}\sum_{j=1}^{n_{grid}} Z_i' V_{\hat{\beta}} Z_j, \tag{17}$$

where n_{grid} is the number of grid lines in each dimension. For a grid width of p pixels, the computational intensity necessary to calculate $\hat{Var}(\hat{\mu}_{Mod})$ is reduced by a factor of approximately p^2.

3.4. Analyses

All four estimators were used to calculate $\hat{\mu}$ and $\hat{Var}(\hat{\mu})$ for AOIs of radius 1 km, 2 km, ..., 15 km centered at each of the 14 AOI center points. In addition, all estimators were used to calculate $\hat{\mu}$ and $\hat{Var}(\hat{\mu})$, for the aggregation of the 14 AOIs of radius 1 km, 2 km, ..., 15 km centered at the 14 AOI center points.

For the stratified estimator, strata were constructed using the logistic model pixel predictions, $\hat{\mu}_i$, rounded to the nearest 0.01. Each pixel was assigned to one of the resulting 101 categories, and the categories were grouped into strata subject to two constraints: (1) categories grouped into the same stratum must be adjacent; e.g., the 0.43 and 0.45 categories cannot be grouped into the same stratum unless the 0.44 category is also included; and (2) each stratum must include at least five subplots. Stratifications featuring two, three, and four optimal strata were constructed by selecting strata boundaries to minimize $\hat{Var}(\hat{\mu}_{Str})$. Legitimate concerns may be raised regarding violations of assumptions resulting from using stratifications based on the sample observations to stratify the same sample units from which the observations are obtained. However, Breidt and Opsomer (2008) concluded that the practical effects were minimal, even for relatively small sample sizes.

For the model-assisted estimator, the logistic model prediction, $\hat{\mu}_i$, of the probability of forest for the ith pixel was considered equivalent to an estimate, $\hat{y}_i$, of proportion forest for the pixel.

For the model-based estimator, issues of bias and computational intensity were investigated. Two approaches to bias assessment were used. Because the issue of bias is closely linked to correct model specification, the first approach focused on assessing the quality of fit of the model to the data at the subplot/pixel level. If the model is correctly specified, a graph of observations versus model predictions for a continuous response variable should feature points that lie along a line with intercept 0 and slope 1. However, because the data used to estimate the model parameters are binary, a slightly modified three-step approach was used: (1) all subplot observation/pixel model prediction pairs, $(y_i,\ \hat{\mu}_i)$, were ordered with respect to $\hat{\mu}_i$; (2) the ordered pairs were grouped into categories of equal numbers of pairs, and the group means of the subplot observations and of the pixel model predictions were calculated; and (3) a graph of the observation means versus the model prediction means was constructed. As for continuous variables, in the absence of model lack of fit a graph of the points defined by the corresponding group means of observations and model predictions should be located along the 1:1 line and have intercept of approximately 0 and slope of approximately 1.

The second approach to bias assessment was areal. Estimates, $\hat{\mu}_{Mod}$, obtained using the model-based estimator were compared to estimates, $\hat{\mu}_{SRS}$, obtained using the simple random sampling estimator. Estimates, $\hat{\mu}_{Mod}$, that are consistently within approximately two SRS standard errors of $\hat{\mu}_{SRS}$, indicate no serious lack of model fit and suggest unbiasedness of the estimator.

The effects of approximating $\hat{Var}(\hat{\mu}_{Mod})$ using a systematic grid were assessed by comparing estimates of standard errors using all pixels in AOIs to estimates obtained using pixels located at the intersections of two-dimensional perpendicular grids of 2-, 5-, and 10-pixel widths.

4. Results and Discussion

4.1. Probability-based estimators

The SRS estimators yielded $\hat{\mu}_{SRS} = 0.7516$ and $SE(\hat{\mu}_{SRS}) = 0.0159$ for the 14 AOIs of radius 15 km considered in aggregate. These estimates served as the standard for comparison for the other three estimators.

For the stratified estimator, optimal stratifications were constructed for the 14 AOIs of radius 15 km in aggregate and then applied to AOIs of smaller radii individually. The stratifications using three and four strata yielded little increase in precision over the stratification with two strata (Table 1). These results are not surprising because few subplots were not either completed forested or completely non-forested in which case a reasonably accurate, simple forest/non-forest classification served as an effective stratification. The utility of a stratification for increasing precision is often evaluated using relative efficiency, calculated as,

$$RE = \frac{\hat{Var}\left(\hat{\mu}_{SRS}\right)}{\hat{Var}\left(\hat{\mu}_{Str}\right)} \tag{18}$$

where values of RE greater than 1.0 indicate increasing effectiveness of the stratification. For this study, the optimal stratification with two strata produced $RE = 2.09$, which can be interpreted as meaning that the sample size would have to be increased by a factor of 2.09 to achieve the same reduction in variance as the stratification achieves. For forest inventory applications, the cost of more than doubling the sampling size would be prohibitive; thus, $RE = 2.09$ is of considerable economic importance.

Estimates of variance obtained using the model-assisted difference estimator were similar to those obtained using the stratified estimator. Although estimates obtained with the model-assisted estimator would generally be expected to be smaller, for this study the result can also be attributed to the fact that the subplot observations were nearly all completely non-forested ($y = 0$) or completely forested ($y = 1$). In particular, because the vast majority of subplots assigned to a stratum are either non-forested or forested, within stratum variances are small and therefore the stratified variance estimate is small. Analogously, because the model predictions are close to the sample observations, their squared differences are also small and therefore the model-assisted variance estimate is also small. Finally, because the stratifications are based on the same model predictions as are used for the difference estimator, similar results are obtained.

Table 1
Results for the stratified estimator.

Stratum	Strata boundaries	Number of subplots	Weight	Mean	SE
Two strata					
1	0.00–0.59	150	0.205	0.1444	0.0099
2	0.60–1.00	581	0.795	0.9083	0.0118
All		731		0.7485	0.0110
Three strata					
1	0.00–0.59	150	0.205	0.1444	0.0099
2	0.60–0.91	259	0.354	0.8471	0.0078
3	0.92–1.00	322	0.440	0.9570	0.0022
All		731		0.7500	0.0108
Four strata					
1	0.00–0.43	119	0.163	0.0963	0.0077
2	0.44–0.59	31	0.042	0.3290	0.0400
3	0.60–0.91	259	0.354	0.8479	0.0078
4	0.92–1.00	322	0.440	0.9570	0.0022
All		731		0.7486	0.0106

Despite generally acceptable results with the model-assisted estimator, a few undesirable results are noted, particularly for small areas. Although bias estimates for the naïve model-assisted estimato Eq. (9) tended to be small for individual AOIs of radius 15 km, ranging from −0.19 to 0.02, the estimates were considerably larger for individual AOIs of radius 5 km, ranging from −0.42 to 0.44. These results can be attributed to small numbers of subplots per small AOI of which a large proportion had poor model predictions. Occasionally, the effects of subtracting the bias estimate (Eq. 9) from the naïve estimate (Eq. 8) were estimates of mean proportion forest for small AOIs greater than 1.0 (Table 2). In addition, all these results must be considered somewhat optimistic because the same data were used to estimate bias as were used to estimate the model parameters.

4.2. Model-based estimator

Two approaches to bias assessment for the model-based estimator were used, one at the pixel level that focused on quality of fit of the model to the data and one at the areal level that focused on comparisons of the SRS and model-based estimates. As for the model-assisted estimator, these assessments may be optimistic because the same data were used for these assessments as were used to estimate the model parameters.

At the pixel level, model predictions corresponding to non-forest subplot observations were generally less than 0.1, although some predictions were considerably larger; for forest observations, predictions were generally greater than 0.8, although some were smaller. Multiple factors may lead to deviations between observations and predictions that contribute to model lack of fit. First, tree density on forest subplots varies considerably, even though the entire subplot is characterized as forest. Thus, completely forested subplots with sparse tree cover may be predicted to have smaller probabilities of forest than forested subplots with dense tree cover. Second, spectral differences for subplots with similar tree densities but different age structures or health conditions may produce different estimates of the probability of forest. Third, subplots with tree cover may be characterized as non-forest if the minimum FIA requirements of a 0.4-ha patch with 36.58-m width and 10% stocking are not satisfied. Fourth, the observation for the smaller subplot may not adequately characterize the larger pixel. Fifth, additional factors such as misregistration between subplot locations and the satellite imagery and subplot disturbance between the subplot observation and image acquisition dates are also possible.

Despite a few large deviations between subplot observations and their corresponding model predictions, the graph of means of central subplot observations versus corresponding means of pixel predictions for the ordered groups suggested no serious lack of model fit. Most pairs of observation and prediction means were located relatively close to the 1:1 line with intercept 0 and slope 1 (Fig. 2). A simple linear regression produced intercept and slope estimates of −0.0020 and 1.0008, respectively.

The second assessment was areal and was based on a comparison of $\hat{\mu}_{Mod}$ and $\hat{\mu}_{SRS}$; $\hat{\mu}_{Mod}$ was not compared to $\hat{\mu}_{Str}$ or $\hat{\mu}_{Dif}$ because of the potential confounding effects of incorporating the same model predictions into both estimators. The rationale for comparing $\hat{\mu}_{Mod}$ and $\hat{\mu}_{SRS}$ is that if

$$\hat{\mu}_{SRS} - 2\sqrt{\hat{Var}\left(\hat{\mu}_{SRS}\right)} \le \hat{\mu}_{Mod} \le \hat{\mu}_{SRS} + 2\sqrt{\hat{Var}\left(\hat{\mu}_{SRS}\right)}, \tag{19}$$

Table 2
Estimates for individual AOIs of small radii.

AOI	Radius = 2 km				Radius = 3 km				Radius = 4 km				Radius = 5 km			
	n	SRS	Dif	Mod	*n*	SRS	Dif	Mod	*n*	SRS	Dif	Mod	*n*	SRS	Dif	Mod
1	2	1.0000	0.9466	0.8973	2	1.0000	0.9499	0.9006	4	1.0000	0.9390	0.8948	5	1.0000	1.0291	0.8990
2	1	1.0000	0.8782	0.8666	3	0.5133	0.4247	0.8449	3	0.5133	0.4074	0.8276	5	0.5080	0.6175	0.8135
3	0	–	–	0.6127	2	1.0000	0.6767	0.6346	4	0.7500	0.6607	0.6325	5	0.6000	0.6744	0.6518
4	0	–	–	0.8544	0	–	–	0.8779	4	1.0000	0.9175	0.8834	5	1.0000	0.9189	0.8888
5	1	1.0000	0.7550	0.4613	3	0.6667	0.4497	0.4995	4	0.5000	0.4529	0.5346	4	0.5000	0.4357	0.5173
6	2	1.0000	0.9642	0.8871	3	1.0000	0.9460	0.8869	3	1.0000	0.9309	0.8718	6	1.0000	0.8859	0.8381
7	0	–	–	0.6872	3	0.6667	0.8430	0.7207	5	0.8000	0.8165	0.7126	6	0.8333	0.8153	0.7124
8	2	0.0000	–	0.6319	2	0.0000	0.6637	0.6637	3	0.0000	0.4720	0.6741	6	0.1667	0.4491	0.6799
9	1	0.0000	0.2521	0.5912	1	0.0000	0.2906	0.6298	4	0.2500	0.4152	0.6431	7	0.2857	0.4108	0.6566
10	1	1.0000	1.4631	0.6333	3	0.6667	0.9313	0.6454	5	0.8000	0.9822	0.6524	8	0.7500	0.7936	0.6503
11	2	1.0000	0.9208	0.8706	2	1.0000	0.8415	0.7913	5	0.8000	0.7313	0.7890	8	0.7500	0.6793	0.7925
12	2	1.0000	1.1140	0.6758	2	1.0000	1.0759	0.6377	2	1.0000	1.0365	0.5984	4	0.7500	0.8416	0.5881
13	1	1.0000	0.8503	0.8401	2	1.0000	0.8792	0.8622	4	1.0000	0.8910	0.8730	5	1.0000	0.9285	0.8801
14	0	–	–	0.8393	2	0.5000	0.4326	0.8311	3	0.6667	0.5639	0.8088	5	0.8000	0.6855	0.8024

then an argument can be made that $\hat{\mu}_{Mod}$ and $\hat{\mu}_{SRS}$ are not statistically significantly different. Further, the test criterion is conservative in the sense that no account is made for the uncertainty in $\hat{\mu}_{Mod}$. Although a naïve approach would be to combine $\hat{Var}(\hat{\mu}_{SRS})$ and $\hat{Var}(\hat{\mu}_{Mod})$ in the criterion, the two variance estimators are based on such different underlying assumptions that the proper approach to combining them is not apparent. The comparisons indicated that the criterion expressed by Eq. (19) is generally satisfied (Fig. 3). Similar results were obtained by McRoberts (2006) using different sets of AOIs of different sizes.

The effects of using data for a two-dimensional, equally-spaced, perpendicular grid to estimate $\hat{Var}(\hat{\mu}_{Mod})$ were evaluated for grid widths of 2, 5, and 10 pixels. Differences in the variance estimates were extremely small, even for AOIs with small radii, and were not disproportionately positive or negative (Table 3). For a grid width of 10 pixels, use of the grid reduces the computational intensity by a factor of 100 and makes estimation of model-based variances feasible for many applications, particularly small area applications.

Finally, although increases in sample sizes produced corresponding decreases in $\hat{Var}(\hat{\mu}_{SRS})$, $\hat{Var}(\hat{\mu}_{Str})$, and $\hat{Var}(\hat{\mu}_{Dif})$, such was not generally the case for $\hat{Var}(\hat{\mu}_{Mod})$ (Table 3). This somewhat counter intuitive phenomenon can be attributed to the fact that $\hat{\mu}_{Mod}$ is a synthetic estimator in the sense that the sample observations used to estimate the model parameters and their covariance matrix come from the entire study area, not just the AOIs. Thus, because $V_{\hat{\beta}}$ is the primary factor in determining $\hat{Var}(\hat{\mu}_{Mod})$, and because $V_{\hat{\beta}}$ does not change for individual AOIs, regardless of their sizes, $\hat{Var}(\hat{\mu}_{Mod})$ remains nearly constant for AOIs of all sizes.

5. Conclusions

Four conclusions may be drawn from this study. First, all four estimators produced similar estimates for mean proportion forest for large AOIs, although only the model-based estimator produced consistently reliable estimates for small AOIs. For small areas, the stratified estimator is less useful than the SRS and model-assisted difference estimators because the former requires minimum numbers of observations for each of multiple strata, not just for the entire AOI. However, for small areas with small sample sizes, the model-assisted difference estimator occasionally produced estimates beyond the bounds of possible values. Second, use of the map in the form of model predictions with the stratified and model-assisted difference estimators enhanced inferences by producing meaningful variance reductions. Third, bias for the model-based estimator was assessed using two approaches, one based on a direct assessment of model lack of fit and one based on comparing estimates obtained using the model-based estimator and estimates obtained using the unbiased SRS estimator. Neither approach indicated model misspecification or bias in the estimator. Fourth, the grid sampling approach for

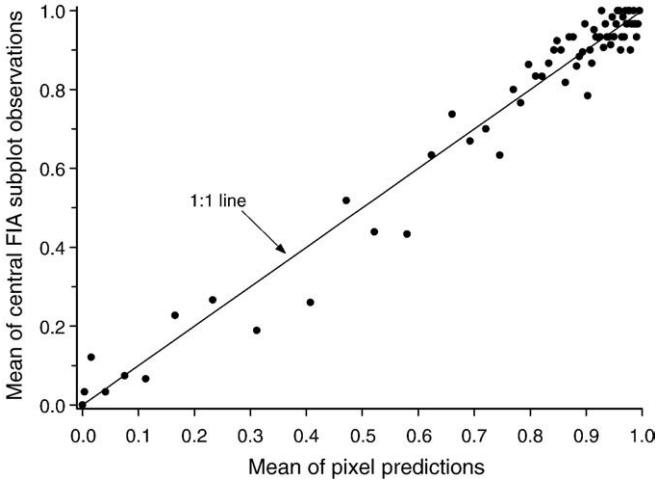

Fig. 2. Group means of central FIA subplot observations versus group means of predictions for pixels containing subplot centers for the 2232 subplots in the study area; group size = 30.

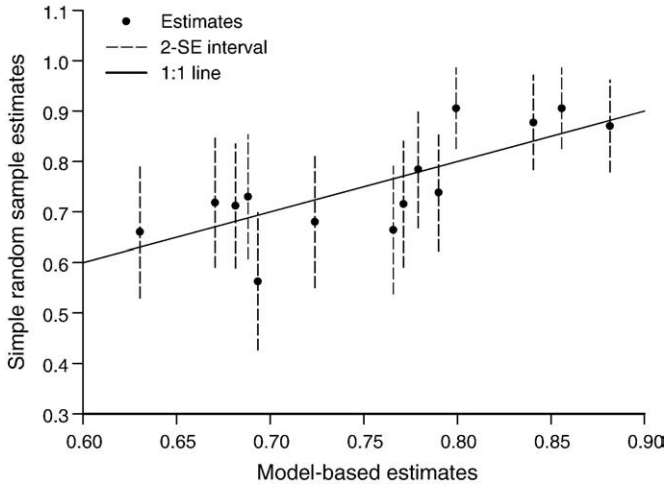

Fig. 3. Simple random sampling estimates versus model-based estimates for circular AOIs of radius 15 km.

Table 3
Estimates for aggregations of 14 AOIs.

Radius[a] (km)	Area (km[b])	Estimator															
		SRS			Stratified[b]						Difference		Model-based				
		n	Mean	SE	Number of strata						Mean	SE	Mean	Grid width (pixels)			
					2		3		4					1	2	5	10
					Mean	SE	Mean	SE	Mean	SE				SE	SE	SE	SE
1	43.98	3	0.3333	0.3333	[c]	[c]	[c]	[c]	[c]	[c]	0.7184	0.1656	0.7544	0.0075	0.0073	0.0079	0.0072
2	175.93	15	0.8000	0.1069	0.8951	0.0428	[c]	[c]	[c]	[c]	0.8749	0.0729	0.7392	0.0071	0.0070	0.0072	0.0063
3	395.84	30	0.7180	0.0818	0.7432	0.0647	0.7474	0.0601	0.7491	0.0604	0.7458	0.0667	0.7448	0.0070	0.0070	0.0071	0.0066
4	703.72	53	0.7272	0.0610	0.7539	0.0428	0.7312	0.0362	0.7299	0.0362	0.7349	0.0450	0.7426	0.0069	0.0069	0.0070	0.0067
5	1099.56	78	0.6992	0.0519	0.7260	0.0373	0.7139	0.0332	0.7153	0.0326	0.7188	0.0385	0.7408	0.0067	0.0067	0.0068	0.0066
6	1583.36	115	0.6569	0.0443	0.6935	0.0316	0.6939	0.0302	0.6974	0.0291	0.7033	0.0306	0.7450	[d]	0.0067	0.0068	0.0067
7	2155.13	158	0.7010	0.0361	0.7102	0.0262	0.7113	0.0256	0.7100	0.0239	0.7112	0.0254	0.7513	[d]	0.0068	0.0068	0.0068
8	2814.86	196	0.7192	0.0318	0.7309	0.0231	0.7295	0.0223	0.7241	0.0206	0.7250	0.0221	0.7556	[d]	0.0068	0.0069	0.0068
9	3562.56	256	0.7420	0.0272	0.7291	0.0200	0.7377	0.0198	0.7351	0.0201	0.7347	0.0193	0.7586	[d]	0.0069	0.0069	0.0070
10	4398.23	319	0.7542	0.0239	0.7396	0.0178	0.7384	0.0171	0.7388	0.0171	0.7405	0.0171	0.7584	[d]	0.0068	0.0069	0.0069
11	5321.85	393	0.7611	0.0213	0.7458	0.0152	0.7433	0.0148	0.7445	0.0148	0.7464	0.0146	0.7516	[d]	[d]	0.0067	0.0067
12	6333.45	466	0.7500	0.0199	0.7462	0.0136	0.7466	0.0134	0.7464	0.0134	0.7459	0.0132	0.7569	[d]	[d]	0.0067	0.0067
13	7433.00	554	0.7551	0.0181	0.7553	0.0127	0.7551	0.0122	0.7576	0.0120	0.7521	0.0121	0.7548	[d]	[d]	0.0066	0.0066
14	8620.52	638	0.7542	0.0169	0.7560	0.0114	0.7587	0.0111	0.7575	0.0109	0.7547	0.0111	0.7548	[d]	[d]	0.0066	0.0066
15	9896.01	731	0.7516	0.0159	0.7485	0.0110	0.7500	0.0108	0.7503	0.0106	0.7488	0.0108	0.7551	[d]	[d]	0.0065	0.0065

[a] Radius of each of the 14 AOIs that are aggregated.
[b] Minimum of five plots per stratum.
[c] Estimates not calculated because of insufficient numbers of plots per stratum.
[d] Estimates not calculated because of computational intensity.

calculating the variance of the model-based estimator produced reductions in computational intensity by factors of 100 with no apparent adverse effects on variance estimates. This result increases the appeal of the model-based estimator and makes it much more feasible for larger AOIs. Finally, although the nature of the data with mostly completely non-forested or completely forested sample observations is not unusual for relatively small inventory plots, moderate caution should be exercised before extrapolating these conclusions to markedly different data sets.

Acknowledgements

The author gratefully acknowledges careful reviews and constructive comments from two reviewers, particularly as they pertained to terminology and assessment of model structure.

Appendix A

$$\hat{\mu}_{\text{Dif}} = \hat{\mu}_{\text{naive}} - \frac{1}{N}E\left[\sum_{i=n+1}^{N}\left(\hat{y}_i - y_i\right)\right]$$
$$= \frac{1}{N}\left(\sum_{i=1}^{n} y_i + \sum_{i=n+1}^{N}\hat{y}_i\right) - \frac{1}{N}(N-n)\left[\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)\right]$$
$$= \frac{1}{N}\left(\sum_{i=1}^{n} y_i + \sum_{i=n+1}^{N}\hat{y}_i\right) - \frac{N-n}{N}\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - \hat{y}\right)$$
$$\approx \frac{1}{N}\left(\sum_{i=1}^{n} y_i + \sum_{i=n+1}^{N}\hat{y}_i\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right) \text{ for } n \ll \text{N}$$
$$= \frac{1}{N}\left(\sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\hat{y}_i + \sum_{i=1}^{n}\hat{y}_i + \sum_{i=n+1}^{N}\hat{y}_i\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)$$
$$= \frac{1}{N}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right) + \frac{1}{N}\sum_{i=1}^{N}\hat{y}_i - \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)$$
$$= \frac{1}{N}\sum_{i=1}^{N}\hat{y}_i - \left(\frac{1}{N} + \frac{1}{n}\right)\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)$$
$$\approx \frac{1}{N}\sum_{i=1}^{N}\hat{y}_i - \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right) \text{ for } N \text{ large and } n \ll N.$$

Appendix B

$$Var(\hat{\mu}_{Dif}) = E(\hat{\mu}_{Dif} - \mu)^2$$
$$= E\left[\frac{1}{N}\left(\sum_{i=1}^{n} y_i + \sum_{i=n+1}^{N}\hat{y}_i\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right) - \mu\right]^2$$
$$= E\left[\frac{1}{N}\sum_{i=1}^{N} y_i + \frac{1}{N}\sum_{i=n+1}^{N}\left(\hat{y}_i - y_i\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right) - \mu\right]^2$$
$$= E\left[\frac{1}{N}\sum_{i=n+1}^{N}\left(\hat{y}_i - y_i\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)\right]^2$$
$$= \frac{1}{N^2}E\left[\sum_{i=n+1}^{N}\left(\hat{y}_i - y_i\right)\right]^2 + \frac{1}{n^2}E\left[\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)\right]^2$$
for independent residuals
$$= \frac{N-n}{N^2}E\left(\hat{y}_i - y_i\right)^2 + \frac{n}{n^2}E\left(\hat{y}_i - y_i\right)^2$$
$$= \left[\left(\frac{1}{N} - \frac{n}{N^2}\right) + \frac{1}{n}\right]\frac{1}{n-1}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2$$
$$\approx \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 \text{ for } N \text{ large and } n \ll N.$$

References

Baffetta, F., Fattorini, L., Franceschi, S., & Corona, P. (2009). Design-based approach to k-nearest neighbours technique for coupling field and remotely sensed data in forest surveys. *Remote Sensing of Environment*, *11*(3), 463−475.

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. Southern Research Station, USDA Forest Service, Asheville, North Carolina. 85 pp. Available at: http://www.srs.fs.usda.gov/pubs/ Last accessed: April 2009.

Breidt, F. J., & Opsomer, J. D. (2008). Endogenous post-stratification in surveys: Classifying with a sample-fitted model. *The Annals of Statistics*, *36*(1), 403−427.

Brewer, K. R. W. (1963). Ratio estimation in finite populations: Some results deductible from the assumption of an underlying stochastic process. *Australian Journal of Statistics*, *5*, 93−105.

Chambers, R. (2003). An introduction to model-based survey sampling. Available at: http://www.eustat.es/prodserv/seminario_i.html Last accessed: November 2009.

Cienciala, E., Tomppo, E., Snorrason, A., Broadmeadow, M., Colin, A., Dunger, K., et al. (2008). Preparing emission reporting from forests: Use of national forest inventories in European countries. *Silva Fennica*, *42*(1), 73−88.
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.
Graubard, B. I., & Korn, E. L. (2002). Inference for superpopulation parameters using sample surveys. *Statistical Science*, *17*(1), 73−96.
Gregoire, T. G. (1998). Design-based and model-based inference: Appreciating the difference. *Canadian Journal of Forest Research*, *28*, 1429−1447.
Hansen, M. H., Madow, W. G., & Tepping, B. J. (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association*, *78*, 776−793.
Kangas, A., & Maltamo, M. (2006). *Forest inventory, methodology and applications.*Dordrecht, The Netherlands: Springer 363 pp.
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. New York: Longman Group, Ltd.
Liu, C., Zhang, L., Davis, C. J., Solomon, D. S., Brann, T. B., & Caldwell, L. E. (2003). Comparison of neural networks and statistical classification of ecological habitats using FIA data. *Forest Science*, *49*(4), 619−631.
Lohr, S. (1999). *Sampling: Design and analysis.*Pacific Grove, CA: Duxbury 494pp.
Magnussen, S., Boudewyn, P., Wulsder, M., & Seeman, D. (2001). Predictions of forest inventory cover type proportions using Landsat TM. *Silva Fennica*, *34*, 351−370.
Mandallaz, D. (2008). *Sampling techniques for forest inventories.*New York: Chapman & Hall 256pp.
Mátern, B. 1960. Spatial variation. Medd. Statens Skogsforskningsinst. Band 49, No. 5. (Reprinted as volume 36 of the series Lecture notes in Statistics. 1986. Springer-Verlag, New York. 151 pp.)
McRoberts, R. E. (2006). A model-based approach to estimating forest area. *Remote Sensing of Environment*, *103*, 56−66.
McRoberts, R. E. (2009). A two-step nearest neighbors algorithm using satellite imagery for predicting forest structure within species composition classes. *Remote Sensing of Environment*, *113*, 532−545.
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. D. (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.
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.
Särndal, C. -E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling.* New York: Springer-Verlag, Inc. 694 pp.
Ministerial Conference on the Protection of Forests in Europe. (2009). Ministerial Conference on the Protection of Forests in Europe. Available at: http://www.mcpfe.org/ Last accessed: March 2009.
Montréal Process. (2005). Montréal Process. Available at: http://www.rinya.maff.go.jp/mpci/ Last accessed: March 2009.
Kyoto Protocol. (1997). The Kyoto Protocol to the framework convention on climate change. Available at: unfccc.int/essential_background/Kyoto_protocol/background/items/1351.php Last accessed: April 2009.
Rennolls, K. (1982). The use of superpopulation-prediction methods in survey analysis, with application to the British National Census of Woodlands and Trees. In H. G. Lund (Ed.), *In place resource inventories: Principles and practices. 9–14 Aug. 1981, Orono,Me* (pp. 395−401). Bethesda, Maryland: Society of American Foresters.
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.
Royall, R. M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, *57*(2), 377−387.
Royall, R. M., & Herson, J. (1973). Robust estimation in finite populations II. *Journal of the American Statistical Association*, *68*(344), 890−893.
Tomppo, E., Korhonen, K. T., Neikinen, J., & Yli-Kojola, H. (2001). Multi-source inventory of the forests of the Hebei Forestry Bureau, Heilongjian, China. *Silva Fennica*, *35*(3), 309−328.
Tomppo, E., Gagliano, C., De Natale, F., Katila, M., & McRoberts, R. E. (2009). Predicting categorical forest variables using an improved k-Nearest Neighbour estimator and Landsat imagery. *Remote Sensing of Environment*, *113*, 500−517.
United Nations Framework Convention on Climate Change (UNFCCC). (2007). National inventory submissions 2007. Available at: http://unfccc.int/national_reports/items/1408.php Last accessed: April 2009.
Valliant, R. (1985). Nonlinear prediction theory and the estimation of proportions in a finite population. *Journal of the American Statistical Association*, *80*(391), 631−641.
Valliant, R., Dorfman, A. H., & Royall, R. M. (2000). *Finite population sampling and inference.* New York: Wiley 504 pp.
Yang, L., Stehman, S. V., Smith, J. H., & Wickham, J. D. (2001). Thematic accuracy of the MLRC land cover for the eastern United States. *Remote Sensing of Environment*, *76*, 418−422.