Remote Sensing of Environment 125 (2012) 157–166

Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

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

Post-stratified estimation of forest area and growing stock volume using lidar-based stratifications

Ronald E. McRoberts [a,*], Terje Gobakken [b], Erik Næsset [b]

[a] Northern Research Station, U.S. Forest Service, 1992 Folwell Avenue, Saint Paul, Minnesota, USA
[b] Department of Ecology and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway

ARTICLE INFO

Article history:
Received 7 March 2012
Received in revised form 24 June 2012
Accepted 7 July 2012
Available online 12 August 2012

Keywords:
National forest inventory
Nonlinear logistic regression model
k-Nearest Neighbors

ABSTRACT

National forest inventories report estimates of parameters related to forest area and growing stock volume for geographic areas ranging in size from municipalities to entire countries. Landsat imagery has been shown to be a source of auxiliary information that can be used with stratified estimation to increase the precision of estimates, although the increase is greater for estimates of forest area than for estimates of growing stock volume. The objective of the study was to assess the utility of lidar-based stratifications for increasing the precision of mean proportion forest area and mean growing stock volume per unit area. Stratifications based on nonlinear logistic regression model predictions of volume obtained from lidar data reduced variances of mean growing stock volume estimates by factors as great as 3.2 and variances of mean proportion forest area estimates by factors as great as 1.5.

Published by Elsevier Inc.

1. Introduction

National forest inventories (NFI) report estimates of parameters related to forest area and growing stock volume for geographic areas such as counties, states, and provinces based on data collected from arrays of field plots. Because of budgetary constraints and natural variability among plots, sufficient numbers of plots frequently cannot be measured to satisfy precision guidelines for the estimates of some parameters unless the estimation process is enhanced using auxiliary information. Remotely sensed data has been shown to be a source of auxiliary information that can be used with stratified estimation to increase the precision of estimates.

1.1. Stratified estimation

Stratified estimation is a statistical technique that can be used to increase the precision of estimates without increasing sample sizes. The essence of stratified estimation is to aggregate observations of the response variable into groups or strata that are more homogeneous than the population as a whole. The population mean and its variance are estimated as weighted means of within-strata means and variances where the weights are based on strata sizes.

If the stratification is accomplished prior to sampling and the within-stratum variances are known or can be easily estimated, then greater precision may be achieved by selecting within-strata sampling intensities to be proportional to within-strata variances (Cochran, 1977). However, NFIs often use permanent plots whose locations are based on systematic grids or tessellations and use sampling intensities that are constant over large geographic areas, if not the entire population. In such cases, even though stratified sampling is not possible, considerable increase in precision may still be achieved simply by using post-sampling stratification, also characterized as post-stratification.

1.2. Applications

Aerial photography served as the earliest source of remotely sensed information for constructing stratifications. With this approach, characterized as double sampling for stratification, an extensive sample of photo plots on aerial photographs is interpreted, photo plots are assigned to strata using ocular methods, and strata weights are estimated as proportions of photo plots assigned to strata. Field crews then visit a subset of the photo plots and observe and measure plot attributes. Plots are assigned to the strata of their corresponding photo plots. Estimates based on these data are then calculated using stratified estimation techniques (Cochran, 1977). Examples of double sampling for stratification using aerial photography are provided by Bickford (1953, 1960), Lawrence and Walker (1954), Kendall and Sayn-Wittgenstein (1961), Macpherson (1962), and Poso (1972), and the topic is addressed in textbooks such as Loetsch and Haller (1964) and Gregoire and Valentine (2008).

More recently, satellite imagery has served as a source of information for constructing stratifications. With this approach, image pixels with centers in the population are classified with respect to land cover attributes, and the classes or aggregations of the classes are then used as strata. Strata weights are calculated as the proportions of pixels in strata, and plots are assigned to strata on the basis of the strata assignments of the pixels containing the plot centers. In Finland, Poso et al. (1984,

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

0034-4257/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.rse.2012.07.002

1987), Poso et al. (1987) derived stratifications from unsupervised classifications of satellite imagery. In the United States of America (USA), Hansen and Wendt (2000) constructed strata by collapsing the classes of the Gap Analysis Program (GAP) classification (Scott et al., 1993) into forest and non-forest strata and then constructing forest edge and non-forest edge strata along forest/non-forest boundaries. Two sets of stratified estimates of forest area and growing stock volume for the states of Illinois and Indiana, USA, were compared: 1986 estimates obtained using a double sampling for stratification approach based on first-phase interpretation of aerial photographs and 1998 estimates using the GAP classification as basis for stratification. Although variances of estimates for the image-based stratification were slightly larger than for the photo-based stratification, the image-based stratification was much more consistent and much less costly to construct.

Also in the USA, McRoberts et al. (2002a) used the 1992 National Land Cover Dataset (NLCD) (Vogelmann et al., 2001) to construct strata similar to those used by Hansen and Wendt (2000). Variances of estimates of mean proportion forest area for the American states of Indiana, Iowa, Minnesota, and Missouri were reduced by factors ranging from 1.72 to 3.22. For two study areas in Minnesota, McRoberts et al. (2002b) used a nearest neighbors approach to predict proportion forest from Landsat imagery with 30-m x 30-m resolution and NFI plot observations of forest/non-forest. Classes of proportion forest predictions were then aggregated into four strata for use in a post-stratification approach. For estimating mean proportion forest area, variances were reduced by factors in the range 4.3–5.6. For study areas in the American states of Minnesota and Wisconsin, McRoberts et al. (2006) used a logistic regression model to predict the probability of forest for Landsat pixels. Classes of the probability of forest were then aggregated into strata for a similar approach to post-stratification. Reductions in variances of estimated mean proportion forest were in the range 3.6–5.9 and reductions for growing stock volume were in the range 1.3–2.5.

Liknes et al. (2004) and Holden et al. (2005) investigated the effectiveness of 250-m x 250-m resolution MODIS-based maps as bases for stratifications. Liknes et al. (2004) focused on stratified estimation of mean proportion forest area and found that although the MODIS-based stratified estimates were more precise than the simple random sampling estimates, they were less precise than Landsat-based stratified estimates. The latter result was attributed to the finer spatial resolution of the Landsat imagery. Holden et al. (2005) used a MODIS-based biomass map as the basis for stratified estimation of biomass for 11 states in the north central region of the USA. The map was only marginally effective at increasing the precision of biomass estimates. This result can be attributed to a weak relationship between plot-level biomass and corresponding MODIS-based predictions. The weakness of the relationship can, in turn, be attributed to the difficulty of predicting below-canopy attributes such as biomass from spectral data that responds primarily to canopy-level attributes. In addition, the spatial size difference between the 62,500-m^2 MODIS pixels and the approximately 672-m^2 plots was likely also a contributing factor.

The important lessons from the Liknes et al. (2004), Holden et al. (2005), and the McRoberts et al. (2002a,b, 2006) studies are that the effectiveness of stratifications based on spectral information from satellite imagery is much greater for canopy-level attributes such as mean proportion forest area than for below-canopy attributes such as mean biomass per unit area and when using imagery whose resolution is closer to the plot size.

Recent reports of strong relationships between below-canopy forest attributes such as growing stock volume and lidar metrics suggest that lidar-based stratifications may be effective for increasing the precision of estimates of parameters related to growing stock volume, biomass, and carbon. For example, Næsset (2002) reported that 80–93% of the variability in field measured volume could be explained by models that use lidar metrics, and Næsset and Gobakken (2008) reported that 88% of the variability in above-ground biomass could be explained with models using lidar metrics. Similar results have also been reported for multiple other studies including Frazer et al. (2011), Li et al. (2008), and Zhao et al. (2009).

1.3. Objectives

The objective of the study was to assess the utility of lidar data as the basis for post-stratifications for increasing the precision of estimates of mean proportion forest area (FOR) and mean growing stock volume (m^3/ha) (VOL).

2. Data

The study area is in Hedmark County, Norway, mostly in Åmot and Stor-Elvdal municipalities (Fig. 1). The study area consists of 2385 km^2 and features altitudinal variations ranging from 204 to 1134 m above sea level (asl) with a mean of 570 m asl. The dominant tree species are Norway spruce (*Picea abies* (L.) Karst.) and Scots pine (*Pinus sylvestris* L.). For the period 1961–1990, mean January and July temperatures were −11° and 13° Celsius, respectively (Norwegian Meteorological Institute, 2012).

2.1. Lidar data

A PA31 Piper Navajo aircraft carried the Optech ALTM 3100 laser scanning system. The laser scanner data were acquired between 15 July 2006 and 12 September 2006 from a height of approximately 1700 m with average speed of 75 ms^{-1}. The pulse repetition frequency was 50 kHz, and the scan frequency was 31 Hz. The maximum scan angle was 16°, which corresponded to an average swath width of approximately 975 m. Pulses transmitted at scan angles that exceeded 14° were excluded from the final dataset. The mean footprint diameter was approximately 50 cm, and the average point density was 0.7 m^{-2}.

The initial processing of the data was accomplished by the contractor (Blom Geomatics, Norway). Planimetric coordinates and ellipsoidal height values were computed for all echoes. Ground echoes were found and classified using the progressive Triangular Irregular Network (TIN) densification algorithm (Axelsson, 2000) of the TerraScan software (Anonymous, 2005). A TIN was created from the planimetric coordinates and corresponding heights of the laser echoes classified as ground points. The ellipsoidal height accuracy of the TIN model was expected to be around 20–30 cm (Kraus & Pfeifer, 1998; Reutebuch et al., 2003). The heights above the ground surface were calculated for all echoes by subtracting the respective TIN heights from the height values of all echoes recorded. The ALTM 3100 sensor is capable of recording up to four echoes per pulse. Data for only single echoes or the first of multiple echoes were used. For each plot and population unit, height distributions were estimated for heights greater than 2 m. Echoes with heights less than 2 m were considered to have been reflected from non-tree objects such as shrubs, grass, or the ground.

The study area was tessellated into population units consisting of square grid cells with the same 250-m^2 area as the field plots (Section 2.2). For each plot and population unit, heights corresponding to the 10th, 20th, …, 100th percentiles of the distributions were denoted $h_1, h_2, \ldots, h_{10}$, respectively. In addition, mean heights and coefficients of variation for the canopy height distributions were calculated. Canopy densities were also calculated as the proportions of echoes with heights greater than 0%, 10%, …, 90% of the 95th height percentile and denoted $d_0, d_1, \ldots, d_9$, respectively (Gobakken & Næsset, 2008).

2.2. Field data

The field measurements were obtained from Norwegian NFI field plots. The Norwegian NFI is a continuous forest inventory system with

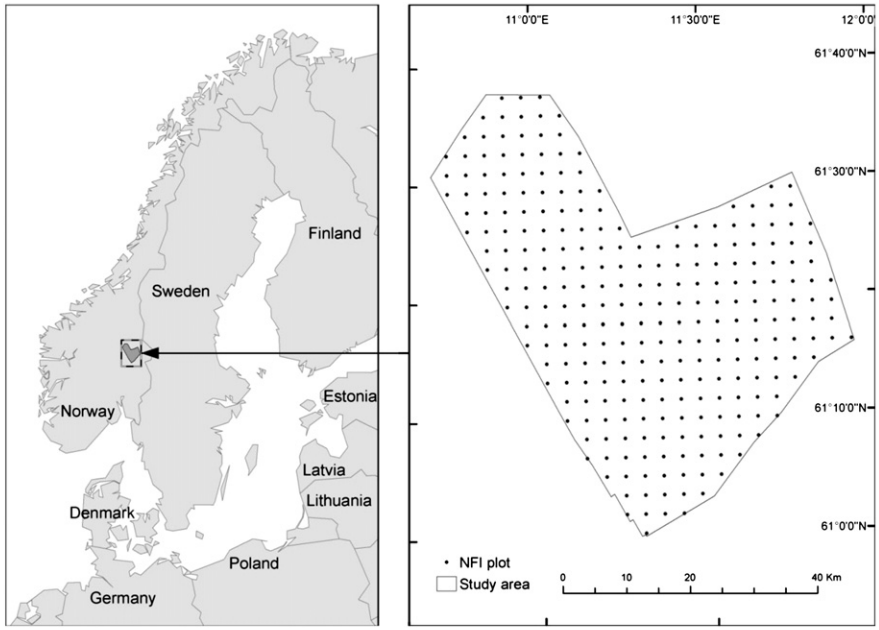

Fig. 1. Study area in Hedmark County, Norway, with illustration of sampling design.

permanent, 250-m^2 circular plots located at the intersections of a 3-km x 3-km grid covering the entire country. The study area includes 266 plots that are remeasured every fifth year following the ordering of a country-level Latin square design with 45-km × 45-km blocks (Tomter et al., 2010).

On each sample plot, all trees with diameter at-breast-height (dbh, 1.3 m) of at least 5 cm were callipered, and tree heights were measured on an average of 10 sample trees per plot selected with probability proportional to stem basal area. Models were used to predict heights for trees whose heights were not measured. (Fitje & Vestjordet, 1977; Vestjordet, 1968). The volume of each sample tree was estimated using species-specific volume models with dbh and measured height or predicted height as predictor variables (Braastad, 1966; Brantseg, 1967; Vestjordet, 1967). The ratio of the mean volume using predicted heights and the mean volume using measured heights was used to adjust the volumes obtained using predicted heights (ratio estimation). The mean and maximum individual tree volumes were 0.117 m^3 and 1.985 m^3, respectively. Total plot volume was obtained by adding volume estimates for individual trees. Although the plot volumes are estimates, the uncertainty of the estimates was considered negligible relative to plot-to-plot variability and was ignored.

Differential Global Navigation Satellite Systems (dGNSS) were used to determine the position of the center of each sample plot. Topcon LegacyE dual-frequency receivers for observing the pseudorange and carrier phase of both the American GPS and the Russian GLONASS were used. Collection of location data lasted a minimum 15 minutes for each plot, with a two-second logging rate. The antenna height ranged from approximately 2 to 4 m. Topcon LegacyE receivers were also used as base stations. The distances between the plots and the base stations were < 50 km. Pinnacle version 1.00 (Anonymous, 1999) post-processing software was used to compute coordinates using the nearest base station as reference. Based on the positional standard errors reported by Pinnacle, the estimated precision of the planimetric plot coordinates ranged from 0 to 2 m, with an average of 0.05 m.

The combination of field measurements for years 2005–2007, which bracket the lidar acquisition dates, and lidar metrics was designated the *calibration dataset* and was used to develop procedures for predicting FOR and VOL. Because of the Latin square sampling design, the field data for years 2005–2007 in the calibration dataset did not represent the entire study area, did not constitute an equal probability sample, and therefore were not used to estimate population parameters. However, the calibration dataset was still sufficient to serve as the basis for developing models and prediction techniques. Field data for years 2006–2010 covered the entire study area and were used to estimate population parameters. The combination of the 2006–2010 field data and the lidar metrics was designated the *estimation dataset*.

3. Methods

This section addresses four issues. First (Section 3.1), a brief overview of stratified estimators and conditions under which stratified estimation is effective is presented. Second (Section 3.2), three techniques used to construct forest attribute maps that serve as bases for stratifications are described: a multiple linear regression model, a nonlinear logistic regression model, and the *k*-Nearest Neighbors (*k*-NN) technique. Third (Section 3.3), issues related to selecting strata boundaries and constructing strata are discussed, and fourth (Section 3.4), a brief statement of how the stratifications were applied is presented.

3.1. Stratified estimators

Cochran (1977) provides a stratified estimator for the population mean:

$$\hat{\mu}_{\text{Str}} = \sum_{h=1}^{H} w_h \hat{\mu}_h, \tag{1}$$

where

$$\hat{\mu}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi}; \tag{2}$$

$h = 1,\ldots H$ denotes strata; y_{hi} is the ith observation in the hth stratum of the variable of interest; w_h is the known weight for the hth stratum, calculated as the proportion of the population in the stratum; n_h is the number of plots assigned to the hth stratum; and $\hat{\mu}_h$ is the sample mean for the hth stratum. For within-strata sample variances calculated as,

$$\hat{\sigma}_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left(y_{hi} - \hat{\mu}_h\right)^2, \tag{3}$$

a stratified estimator of the variance of the estimated mean for fixed strata sample sizes is,

$$\hat{\text{Var}}\left(\hat{\mu}_{\text{Str}}\right) = \sum_{h=1}^{H} w_h^2 \frac{\hat{\sigma}_h^2}{n_h}. \tag{4a}$$

For random within-strata sample sizes, such as is usually the case for post-stratification, (Cochran, 1977, p. 135) provides a modified stratified estimator of the variance,

$$\hat{\text{Var}}\left(\hat{\mu}_{\text{Str}}\right) = \sum_{h=1}^{H} w_h \frac{n_h}{n} \frac{\hat{\sigma}_h^2}{n_h} + \frac{1}{n} \sum_{h=1}^{H} (1 - w_h) \frac{n_h}{n} \frac{\hat{\sigma}_h^2}{n_h}, \tag{4b}$$

where n is the total sample size over all strata. For large sample sizes, particularly when the strata weights are nearly equal to the proportion of plots assigned to strata, Eqs. (4a) and (4b) produce similar estimates.

The stratified estimator is unbiased in the sense that over all possible samples $E(\hat{\mu}_{Str}) = \mu$ where E(.) denotes statistical expectation and μ is the true population mean. However, even though the estimator $\hat{\mu}_{Str}$ is unbiased, the estimate obtained with any particular sample may still deviate substantially from μ. The effectiveness of a stratification in reducing variance is often assessed using relative efficiency calculated as,

$$\text{RE} = \frac{\hat{\text{Var}}\left(\hat{\mu}_{\text{SRS}}\right)}{\hat{\text{Var}}\left(\hat{\mu}_{\text{Str}}\right)}, \tag{5}$$

where for sample size, n,

$$\hat{\mu}_{\text{SRS}} = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{6}$$

and

$$\hat{\text{Var}}\left(\hat{\mu}_{\text{SRS}}\right) = \frac{1}{n-1} \sum_{i=1}^{n} \left(y_i - \hat{\mu}_{\text{SRS}}\right)^2 \tag{7}$$

are the simple random sampling (SRS) estimators. Values of RE greater than 1.0 indicate that the stratification has been effective in reducing the variance and increasing precision. RE may also be interpreted as the factor by which the sample size would have to be increased to obtain the same precision with the SRS estimators as would be achieved with the stratified estimators.

Stratified estimation is effective when the elements of a heterogeneous population are grouped into strata so variances of within-strata means are substantially smaller than the variance of the overall mean obtained under the SRS assumption or when strata with large variances for means are small in size. Stratified estimation may be ineffective if the population is partitioned into strata that have similar means and variances or when large numbers of strata reduce within-strata sample sizes and thereby increase variance estimates of within-strata means. Cochran (1977) indicates that there is little additional benefit when more than six to eight strata are used. The Forest Inventory and Analysis (FIA) program of U.S. Forest Service, which conducts the NFI of the USA, has found post-stratification effective for as few as 2 to 4 strata (Hansen & Wendt, 2000; McRoberts et al., 2002a, 2002b, 2006).

When stratifications are constructed using remotely sensed data, conditions leading to increased precision include stratification variables that are closely related to response variables, sampling units that are accurately registered to the remotely sensed data, close proximity of the dates of the remotely sensed data and the dates of the plot measurements, and small errors associated with the stratification. Numerous factors contribute to stratification errors: (i) errors in the underlying remotely sensed data, (ii) errors in the registration of plot locations to the remotely sensed data (McRoberts, 2010), (iii) mixed land covers for spatial units of the remotely sensed data, (iv) diffusion of spectral information among adjacent spatial units of the remotely sensed data, and (v) differences in sizes of plots and spatial units of the remotely sensed data. Although stratifications are more effective at reducing variance when errors in the prediction of the stratification variable from the remotely sensed data are small, larger errors do not induce bias into the stratified estimator but rather only decrease the effectiveness of the stratification for reducing variances.

3.2. Constructing a map of growing stock volume

The first task was to construct a map of VOL using the NFI plot observations and lidar metrics. Three approaches for predicting VOL were investigated: (1) a multiple linear regression model, (2) a nonlinear logistic regression model, and (3) the k-NN technique. Use of a multiple linear regression model to predict VOL from lidar metrics is intuitive and relatively easy. The potential disadvantage is that the range of values for the lidar metrics may be greater for the population as a whole than for the sample. If so, extrapolations resulting from application of the model to values of lidar metrics not observed in the sample may be negative or unrealistically large.

The nonlinear logistic regression model has the mathematical form,

$$y_i = \frac{\alpha}{1 + \exp\left(\beta_0 + \sum_{j=1}^{J} \beta_j x_{ij}\right)} + \varepsilon_i, \tag{8}$$

where i indexes population units, x_{ij} is observation of the jth lidar metric for the ith population unit, α and β_j, $j = 0, \ldots, J$, are parameters to be estimated, and ε_i is a residual error term. This logistic regression model, whose parameters are estimated using least squares techniques, is not to be confused with the logistic regression model used with categorical data (Agresti, 2007). The advantage of the logistic model expressed by Eq. (8) is that all predictions are non-negative and are constrained by the horizontal asymptote, α, which is estimated from the sample data.

For inventory applications, stratified estimation is often used to enhance the precision of estimates of parameters of multiple forest attributes, typically at least mean FOR and mean VOL. To ensure consistency, the same stratification is preferred for estimation of parameters for both attributes. Therefore, a map that is constructed to depict both attributes simultaneously may produce better stratification results than a map constructed for either attribute individually.

The non-parametric, multivariate k-NN technique has enjoyed popularity within the forest inventory community for predicting forest attributes from remotely sensed data (Tomppo et al., 2008; McRoberts et al., 2010; McRoberts, 2012). With this technique, population unit predictions are calculated as linear combinations of observations for the population units in a sample that are most similar, or nearest, in a space of auxiliary variables to the unit requiring a prediction. The k-NN approach has two advantages. First, k-NN can be used to predict a multivariate suite of response variables. Second, VOL predictions cannot exceed the range of observations in the sample. Thus, a stratification based on a multivariate map constructed to optimize prediction of FOR and VOL simultaneously may produce stratified estimates of both means whose precision is jointly optimal. McRoberts et al. (2002b) illustrates use of the k-NN technique in support of stratified estimation.

Let $\mathbf{Y}$ denote a possibly multivariate vector of response variables with observations for a sample of size n from a finite population of size N, and let $\mathbf{X}$ denote a vector of auxiliary variables with observations for all population units. In the terminology of nearest neighbors techniques, the auxiliary variables are designated *feature variables* and the space defined by the feature variables is designated the *feature space*; the set of sample population units for which observations of both response and feature variables are available is designated the *reference set*; and the set of population units for which predictions of response variables are desired is designated the *target set*. All elements of both the reference and target sets are assumed to have a complete set of observations for all feature variables. For continuous response variables, the nearest neighbors prediction, $\hat{y}_i$, for the ith target set element is calculated as,

$$\hat{y}_i = \sum_{j=1}^{k} w_{ij} y_j^i \tag{9}$$

where $\{y_j^i, j=1, 2, \ldots, k\}$ is the set of response variable observations for the k reference set elements that are nearest or most similar to the ith target set element in feature space with respect to a distance metric, d, and w_{ij} is the weight assigned to the jth nearest neighbor with $\sum_{j=1}^{k} w_{ij} = 1$. For this application, the Euclidean distance metric and equal weighting of neighbors were used, and for each combination of feature variables k was selected to minimize the sum of squared residuals,

$$\mathrm{SS}_{\mathrm{res}} = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2, \tag{10}$$

which was calculated using a leave-one-out technique with the reference dataset.

For each of the linear, nonlinear, and k-NN prediction approaches an optimal combination of lidar metrics as predictor variables was sought. Although graphs of observations versus predictions were also used, the primary optimality criterion was model efficiency (Vanclay & Skovsgaard, 1997) calculated as,

$$Q^2 = 1 - \frac{\mathrm{SS}_{\mathrm{res}}}{\mathrm{SS}_{\mathrm{mean}}}, \tag{11}$$

where

$$\mathrm{SS}_{\mathrm{mean}} = \sum_{i=1}^{n} (y_i - \bar{y})^2. \tag{12}$$

Although in some cases Q^2 is equivalent to the more familiar R^2 used for linear regression models, such is not necessarily the case for the nonlinear and k-NN approaches (Anderson-Sprecher, 1994). For the linear regression and k-NN approaches, automated evaluation of all combinations of lidar metrics was easily implemented. For the nonlinear logistic regression model, an iterative, stepwise approach was used whereby the single lidar metric with the greatest value of Q^2 was first selected. Next, all combinations of the first lidar metric with one of the remaining lidar metrics were evaluated, and the combination with the greatest value of Q^2 was selected. This procedure of selecting the lidar metric that best augmented the previously selected metrics continued until a Q^2 value was obtained for each possible number of metrics. For the linear and nonlinear regression models, final selection of predictor variables was based on a subjective assessment of the combination beyond which only negligible increases in Q^2 were found. Unlike regression approaches, inclusion of predictor or feature variables that are unrelated to the response variable may cause Q^2 for the k-NN technique to decrease (McRoberts et al., 2002b). Thus, for the k-NN approach, the combination of feature variables that produced the greatest Q^2 was finally selected.

3.3. Constructing strata

The second task was to construct classes of volume predictions that could be aggregated to form strata. For each prediction approach, the attribute prediction, $\hat{v}_i$, for the ith population unit was standardized by scaling it to the [0,100] interval:

$$s_i = 100 \cdot \frac{\hat{v}_i - \hat{v}_{\min}}{\hat{v}_{\max} - \hat{v}_{\min}}, \tag{13}$$

where $\hat{v}_{\min}$ and $\hat{v}_{\max}$ are the minimum and maximum predictions, respectively, over all population units. The ith unit was then assigned to one of the 101 standardized classes [0,0], (0,1], …, (99,100], and the proportion of units assigned to each class was calculated.

Stratifications were subject to a minimum number of plots per stratum to ensure reliable estimates of within-strata means and variances. Särndal et al. (1992, p. 267, 407) recommend a minimum of either 10 or 20 plots per stratum; Kish (1965) and Cochran (1977, p. 134) recommend a minimum of 10 plots per stratum; and Scott et al. (2005, page 46) recommend a surprisingly small four plots per stratum. Westfall et al. (2011) used American NFI plot data similar to data used for this study to investigate the effects of small sample sizes and reported two conclusions. First, small numbers of plots per stratum produced erroneously small standard errors, and second, 10 plots per stratum were sufficient to produce stable estimates of means and standard errors. For this study, a minimum of 10 plots per stratum was imposed.

Stratifications were constructed using $m = 1, \ldots, 6$ strata. The selection of $m = 6$ as the maximum number of strata was based on multiple factors: (i) Cochran's (1977) recommendation that more than 6–8 strata produce little additional gain in precision, (ii) the experience of McRoberts et al. (2002a,b, 2006) that approximately four Landsat-based strata were optimal, (iii) preliminary analyses for this study that indicated little gain in precision was achieved for more than six strata, (iv) frequent failure to satisfy the minimum number of 10 plots per stratum with more than six strata, and (v) the computational intensity associated with finding optimal strata boundaries for more than six strata.

Stratifications were constructed by aggregating adjacent standardized classes using three approaches, each satisfying a different criterion. First, adjacent classes were aggregated such that the proportion of population units assigned to each stratum was as close as possible to $\frac{1}{m}$ where m is the number of strata. For example, for $m = 3$, the first stratum may have consisted of classes [0,0]–(5,6] whose proportion of population units was 0.36; the second stratum may have consisted of classes (6,7]–(90,91] whose proportion of units was 0.33; and the third stratum may have consisted of classes (91,92]–(99,100] whose proportion of units was 0.31. Because of the non-uniform grouping of population units into classes, proportions of units aggregated into strata did not always equal exactly $\frac{1}{m}$. Second, adjacent classes were aggregated in such a manner that $\mathrm{RE}_{\mathrm{VOL}}$ was maximized for each number of strata. Third, adjacent classes were aggregated in such a manner that $\mathrm{RE}_{\mathrm{FOR}} + \mathrm{RE}_{\mathrm{VOL}}$ was maximized for each number of strata.

Breidt and Opsomer (2008) coined the term *endogenous post-stratification* to describe approaches to post-stratification for which the range of predictions of a stratification variable obtained using observations of the response variable is divided into pre-determined intervals. They also noted that the approach violates two standard stratification assumptions. First, the assumption that plots are assigned to strata without error is violated because of uncertainty associated with the predictions. Second, the assumption that observations are assigned to strata independently of observations of the response variable is violated because the predictions depend on those observations. However, using data similar to those used for this study, Breidt and Opsomer (2008) also demonstrated that when strata are constructed by dividing the range of predictions obtained from a linear model calibrated using the response variable observations, the detrimental effects of violating the stratification assumptions are negligible, even for small sample sizes. Dahlke et al. (2012) also used similar data to show that when a non-parametric prediction procedure was used, estimates of standard errors were asymptotically similar to those obtained using the familiar Horvitz–Thompson estimator (Horvitz & Thompson, 1952). The statistical consistency of results obtained using endogenous post-stratification was confirmed empirically using simulations similar to those reported by Dahlke et al. (2012).

3.4. Application of estimators

The stratified estimators expressed by Eqs. (1) and (4b) were used to calculate estimates of mean FOR and mean VOL for 2–6 strata for stratifications based on each prediction procedure. RE_{FOR} and RE_{VOL} were calculated using Eq. (5), and the sum, $RE_{FOR}+RE_{VOL}$ was then also calculated. Each plot was assigned to the stratum of the population unit containing the plot center. For this application, finite population correction factors were assumed to be negligible and were ignored.

4. Results and discussion

4.1. Predicting volume (VOL) and forest/non-forest (FOR)

4.1.1. Multiple linear regression model

For the linear regression model, the positive effect on Q^2 of including an intercept term was negligible. Thus, further analyses used only the linear regression model without an intercept term. In general, little improvement in Q^2 values was obtained when more than the five best independent variables were used. The five selected independent variables for the linear regression model were h_1, h_3, h_6, h_7, and d_9 and produced $Q^2_{VOL}=0.772$. Graphs of observations versus predictions indicated only modest lack of fit for the linear regression model (Fig. 2). The proportion of population units with values for the selected predictor variables outside the ranges observed for the calibration dataset was 0.020; the proportion of population units with VOL predictions outside the range predicted for the field reference plots was 0.023; and the proportion of estimation plots with VOL predictions outside the range predicted for the calibration plots was 0.034. The extrapolated predictions were sometimes unrealistic with a range of -206.33 m^3/ha to 1442.10 m^3/ha. Although extrapolations were rare, their effects were large. The range of standardized VOL predictions calculated using Eq. (13) for all population units is necessarily 0–100, but the range for the population units containing estimation plot centers was only 9–34 (Fig. 3). This result has the potential to decrease the effectiveness of resulting stratifications because the plot observations are concentrated in a small range of standardized VOL predictions. This concentration reduces the number of possibilities for aggregating classes into strata and thereby inhibits optimal selection of strata boundaries. Use of class widths smaller than 1.0 may alleviate this condition, but then selection of optimal strata boundaries would entail much greater computational intensity. Of importance, the phenomenon cannot be attributed to standardization of the volume predictions but rather to the results of extrapolating the model beyond the range of the data used to estimate its coefficients.

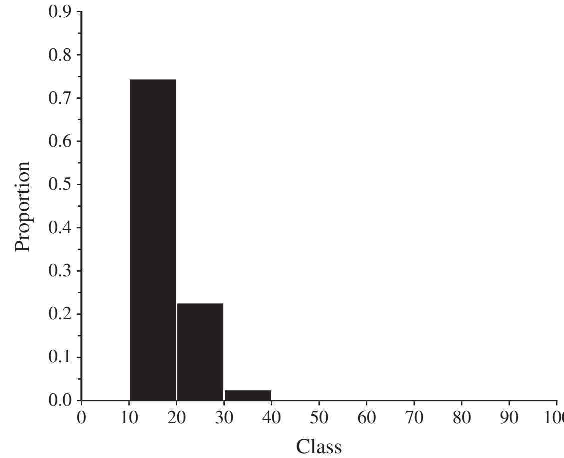

Fig. 3. Distribution of standardized linear model volume class predictions for population units.

4.1.2. Nonlinear logistic regression model

Variables selected for the nonlinear logistic model, in order, were h_1, h_{10}, d_0, and d_9 and produced $Q^2_{VOL}=0.898$, the largest for the three prediction methods. Graphs of VOL observations versus predictions indicated no apparent lack of fit, although one possible outlier was observed (Fig. 4). Serious extrapolation problems are impossible with the nonlinear regression technique because predictions cannot extend below 0 or greater than the estimated asymptote which is obtained from the sample data. For the nonlinear model, the proportion of population units with values for the selected predictor variables outside the range observed for the calibration dataset was 0.026; the proportion of population units with VOL predictions outside the range predicted for the calibration plots was 0.007; and the proportion of estimation plots with VOL predictions outside the range predicted for the calibration plots was 0.064. The range of standardized VOL

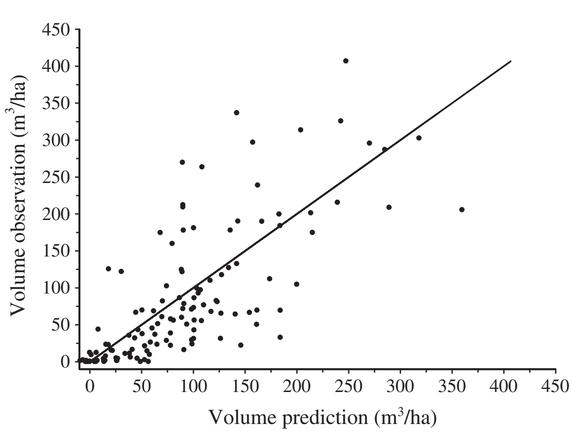

Fig. 2. Plot volume observations versus linear model predictions.

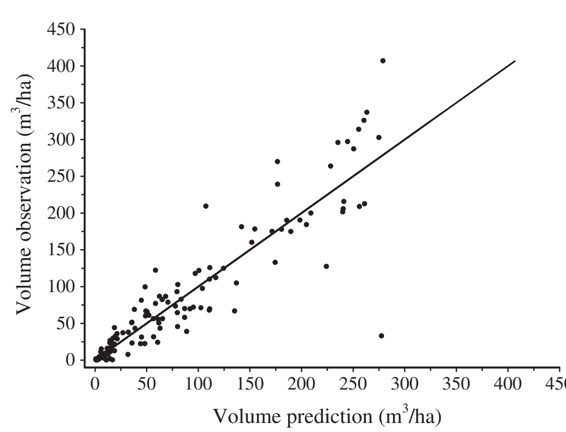

Fig. 4. Plot volume observations versus nonlinear logistic model predictions.

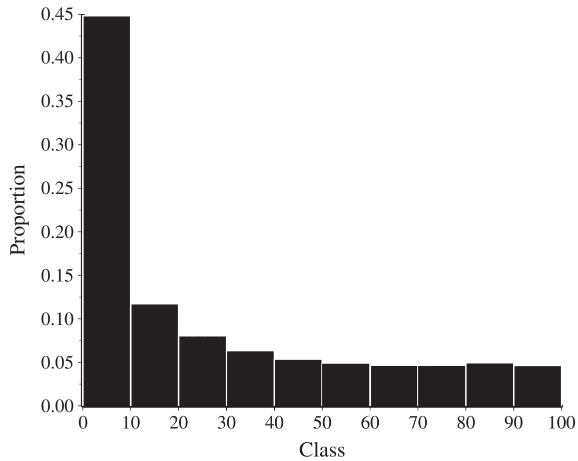

Fig. 5. Distribution of standardized nonlinear logistic model volume class predictions for population units.

predictions calculated using Eq. (13) for population units containing estimation plot centers was 0–100 (Fig. 5).

4.1.3. k-Nearest Neighbors technique

The selected feature variables for the k-NN technique were h_{10}, d_0, d_2, d_3, and d_4 with $k=15$ and produced $Q^2_{FOR}=0.650$and $Q^2_{VOL}=0.714$. Although $Q^2_{VOL}$was smallest for the three prediction methods, the k-NN parameters were selected to maximize the sum, $Q^2_{RE}+Q^2_{VOL}$, not just Q^2_{VOL}. Graphs of observations versus predictions indicated no apparent lack of fit, although several possible outliers were observed (Fig. 6). No extrapolation problems were experienced with the k-NN technique because the predictions cannot extend beyond the range of response variable observations in the reference set. The proportion of population units with values for the selected feature variables outside the range observed for the calibration dataset was 0.024, and the range of standardized k-NN VOL predictions for population units with estimation plots was 0–100 (Fig. 7).

4.2. Stratified estimation

The precision of the stratified estimates of mean FOR and mean VOL were affected by three factors: (1) the prediction procedure of which three approaches were considered, a multivariate linear regression model, a nonlinear logistic regression model, and the k-NN technique; (2) the number of strata which ranged from 1 to 6; and (3) criteria for selecting strata boundaries of which three were considered, equal size strata, maximization of RE_{VOL}, and maximization

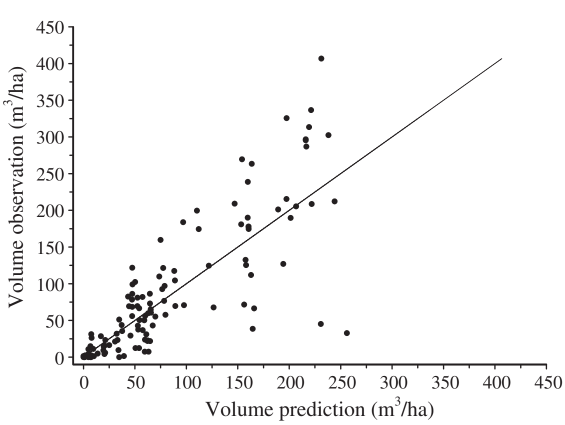

Fig. 6. Plot volume observations versus k-NN predictions.

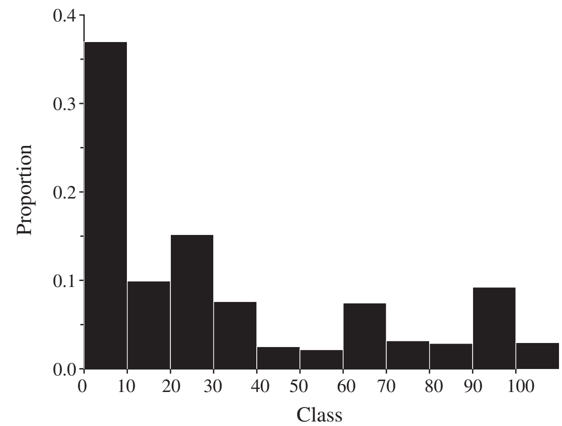

Fig. 7. Distribution of standardized k-NN volume class predictions for population units.

of $RE_{FOR}+RE_{VOL}$. The effects of levels of the three factors were evaluated for three measures: RE_{FOR}, RE_{VOL}, and $RE_{FOR}+RE_{VOL}$. Care must be taken to distinguish between RE_{VOL}, and $RE_{FOR}+RE_{VOL}$ as measures to be maximized when selecting strata boundaries, and RE_{FOR}, RE_{VOL}, and $RE_{FOR}+RE_{VOL}$ as stratification outcomes. Values of the three REs for all levels of the three factors are reported in Table 1.

4.2.1. Modeling procedure

Over all combinations of strata sizes and criteria for selection of strata boundaries, stratifications obtained using the linear model were, with only a few exceptions, inferior with respect to all three optimization criteria. This phenomenon can be attributed to the concentration of plots into a small number of the classes defined by Eq. (13). As a result of the inferior linear model results, this prediction approach is not considered further.

4.2.2. Number of strata

Comparisons of stratifications based on the nonlinear logistic model predictions and the k-NN predictions with respect to the three RE measures produced mixed results. For RE_{VOL} and $RE_{FOR}+RE_{VOL}$, the nonlinear logistic regression model was superior or comparable to the k-NN procedure. However, for RE_{FOR}, the k-NN procedure was superior when using the $RE_{FOR}+RE_{VOL}$ optimization criterion, a result that can be attributed to considering prediction of both FOR and VOL when selecting the k-NN feature variables.

For both the nonlinear logistic model and the k-NN technique, values of RE generally increased for larger number of strata. The increases were nearly always greater for RE_{VOL} and $RE_{FOR}+RE_{VOL}$ than for RE_{FOR}. For the nonlinear logistic model, ratios of RE_{VOL} for six strata to RE_{VOL} for four strata and ratios of $RE_{FOR}+RE_{VOL}$ for six strata to $RE_{FOR}+RE_{VOL}$ for four strata ranged between 1.09 and 1.22. For RE_{FOR}, ratios were only slightly greater than 1.00. For the k-NN technique, ratios for all three RE measures ranged from 1.00 to approximately 1.16. Comparisons of particular values of RE indicated that six strata were superior for stratifications based on nonlinear logistic model predictions. For the k-NN predictions, six strata were only marginally superior. Based on these results, further analyses were restricted to 4–6 strata for stratifications based on both the nonlinear logistic model and k-NN predictions.

4.2.3. Criteria for selecting strata boundaries

For stratifications based on nonlinear logistic model predictions with 4–6 strata, selection of strata boundaries to maximize RE_{VOL} increased RE_{VOL} by multiplicative factors of 1.30 relative to values for equal size strata but produced slight reductions in RE_{FOR}. Selection of strata boundaries to maximize $RE_{FOR}+RE_{VOL}$ produced slightly smaller multiplicative increases in RE_{VOL} and $RE_{FOR}+RE_{VOL}$ but small increases, rather than decreases, in RE_{FOR}.

Table 1
Summary of stratification results.*

Number of strata	Linear			Nonlinear			k-NN		
	RE_{FOR}	RE_{VOL}	$RE_{FOR}+RE_{VOL}$	RE_{FOR}	RE_{VOL}	$RE_{FOR}+RE_{VOL}$	RE_{FOR}	RE_{VOL}	$RE_{FOR}+RE_{VOL}$
Equal strata									
1**	1.000	1.000	2.000	1.000	1.000	2.000	1.000	1.000	2.000
2	1.176	1.462	2.638	1.237	1.472	2.709	1.264	1.353	2.618
3	1.192	1.629	2.721	1.500	1.743	3.243	1.620	1.590	3.210
4	1.192	1.935	3.127	1.489	2.053	3.542	1.719	1.712	3.431
5	1.179	2.034	3.213	1.465	2.262	3.727	1.920	1.691	3.610
6	1.176	2.130	3.306	1.519	2.333	3.852	2.097	1.750	3.847
Strata boundaries selected to maximize RE_{VOL}									
1**	1.000	1.000	2.000	1.000	1.000	2.000	1.000	1.000	2.000
2	1.045	1.792	2.837	1.044	1.980	3.024	1.075	1.589	2.664
3	1.177	2.188	3.365	1.444	2.351	3.796	1.365	1.792	3.158
4	1.177	2.272	3.449	1.466	2.630	4.096	1.366	1.813	3.200
5	1.193	2.336	3.530	1.461	3.024	4.485	1.364	1.848	3.212
6	1.197	2.373	3.570	1.470	3.188	4.657	1.582	1.857	3.439
Strata boundaries selected to maximize $RE_{FOR}+RE_{VOL}$									
1**	1.000	1.000	2.000	1.000	1.000	2.000	1.000	1.000	2.000
2	1.045	1.792	2.837	1.044	1.980	3.024	2.354	1.017	3.371
3	1.177	2.188	3.365	1.489	2.325	3.814	2.347	1.607	3.954
4	1.184	2.270	3.454	1.512	2.601	4.113	2.347	1.750	4.097
5	1.193	2.336	3.530	1.506	2.985	4.491	2.387	1.783	4.170
6	1.197	2.373	3.570	1.514	3.174	4.688	2.448	1.754	4.202

*$RE = \frac{\hat{Var}(\hat{\mu}_{SRS})}{\hat{Var}(\hat{\mu}_{Str})}$.
**Estimates for single stratum equal simple random sampling estimates.

For stratifications based on k-NN predictions with 4–6 strata, selection of strata boundaries to maximize RE_{VOL} increased RE_{VOL} by multiplicative factors of 1.06 to 1.09 relative to values for equal size strata. However, RE_{FOR} decreased by factors of 0.21 to 0.25. Selection of strata boundaries to maximize $RE_{FOR}+RE_{VOL}$ produced only slightly different values for RE_{VOL}, but increases in RE_{FOR}, ranging from 1.17 to 1.37.

For stratifications based on nonlinear model predictions, values for all three RE measures were comparable regardless of whether strata boundaries were selected to maximize RE_{VOL} or $RE_{FOR}+RE_{VOL}$. However, for stratifications based on k-NN predictions, values of RE_{FOR} were substantially greater, but values of RE_{VOL} were substantially less, when strata boundaries were selected to maximize $RE_{FOR}+RE_{VOL}$ rather than RE_{VOL}.

4.3. Stratification summary

Overall, stratifications based on nonlinear logistic regression model predictions and featuring 4–6 strata were superior. Values for all three RE measures were generally comparable regardless of whether strata boundaries were selected to maximize RE_{VOL} or $RE_{FOR}+RE_{VOL}$. However, use of $RE_{FOR}+RE_{VOL}$ as the optimization criterion when selecting strata boundaries permits greater flexibility. For example, values of RE_{VOL} were often approximately twice as large as values of RE_{FOR} when the strata boundary selection criterion was the equally weighted sum of RE_{FOR} and RE_{VOL}. However, different weightings could be used to produce values of RE_{FOR} and RE_{VOL} that are more similar. In summary, the results of this study indicated the superiority of the nonlinear logistic regression model predictions as the basis for 4–6 optimized strata. For equal weighting of precision for mean FOR and mean VOL, RE_{VOL} may be preferable as an optimization criterion because it is slightly less complex, but if flexibility in weighting is desired, a weighted sum of RE_{FOR} and RE_{VOL} may be preferable.

Details of stratifications using the nonlinear logistic model predictions as the basis for the strata and the RE_{VOL} optimization criterion are reported in Tables 2a–2c. For all three stratifications, lower numbered strata represent population units with smaller predicted VOL. The positive relationship between the strata numbers and the strata means indicates that estimation plots with large VOL observations were assigned to population units with large VOL predictions and reflects the strong relationship between the calibration plot observations and the nonlinear logistic model predictions. The stratum-level mean VOL estimates and their standard errors indicate that the stratifications produced the desired effects. In particular, strata with larger standard errors have smaller weights. This feature of stratified estimation whereby groups of plots whose observations are more variable are assigned to smaller strata is a primary reason that stratified estimation reduces overall variance estimates. Of interest, for the 5- and 6-strata cases, the mean for the last stratum is smaller than the means for the two preceding strata. However, the standard error for the last stratum is larger than for the preceding strata. Thus, the explanation for the smaller mean can likely be attributed to use of a criterion based on maximizing REs which are directly related to variances and standard errors but not to means.

For stratum-level mean FOR estimates, all three stratifications reflect the fact that all plots were either completely forested or completely non-forested. For each stratification, mean FOR for all except the first stratum was either 1.000 or very close to 1.000 (Tables 2a,2b,2c). An easy calculation shows that for the latter strata with mean FOR < 1.000, either one or two estimation plots had FOR = 0.000 rather than FOR = 1.000. This situation is likely due to misclassification of the one or the two plots resulting from harvest between the lidar acquisition and plot measurement dates, failure of the plots to adequately characterize the

Table 2a
Estimates for stratifications based on nonlinear logistic model predictions for four strata with strata boundaries selected to maximize RE_{VOL}.

Stratum	Stratum weight	Number of plots	Proportion of plots	Proportion forest area (FOR)		Growing stock volume (VOL)	
				Mean	SE	Mean	SE
1	0.397	122	0.458	0.607	0.044	13.242	1.818
2	0.420	95	0.357	0.989	0.011	80.757	5.509
3	0.116	25	0.094	1.000	0.000	205.909	12.563
4	0.067	24	0.090	0.917	0.058	246.605	34.127
Total	1.000	266	1.000	0.834	0.020	79.589	3.788

Table 2b
Estimates for stratifications based on nonlinear logistic model predictions for five strata with strata boundaries selected to maximize RE_{VOL}.

Stratum	Stratum weight	Number of plots	Proportion of plots	Proportion forest area (FOR)		Growing stock volume (VOL)	
				Mean	SE	Mean	SE
1	0.397	122	0.458	0.607	0.044	13.242	1.818
2	0.420	95	0.357	0.989	0.011	80.757	5.509
3	0.116	25	0.094	1.000	0.000	205.909	12.563
4	0.047	13	0.049	0.923	0.077	332.975	39.541
5	0.020	11	0.041	0.909	0.091	144.531	41.276
Total	1.000	266	1.000	0.834	0.020	81.654	3.533

population units containing the plot centers, or plot location errors. The phenomenon with all strata after the first having mean FOR = 1.000 or very close means that little gain in RE_{FOR} can be realized beyond 2–3 strata.

The stratified population estimates are similar to but not exactly the same as the SRS estimates. For this study, the SRS mean FOR was 0.808 with SE of 0.024 and SRS mean VOL was 76.502 with SE of 6.142. Deviations of stratified estimates of means from the SRS estimates is due to differences between strata weights and the proportions of plots assigned to strata (Tables 2a,2b,2c). Causes for the differences between weights and proportions of plots by strata vary but include a non-representative sample and classification errors. For all cases, deviations in estimates of means were less in absolute value than twice the stratified standard errors.

For this study, RE_{VOL} for the best stratifications was approximately in the range 2.6–3.2. Such values of RE_{VOL} are substantially greater than the 1.3–2.5 values obtained by McRoberts et al. (2006) in the USA using similar NFI plot observations and stratifications based on predictions of FOR obtained from Landsat data. This result partially reflects the difference between predictions of FOR and predictions of VOL, but it also reflects the greater utility of lidar data for characterizing sub-canopy forest attributes. For similar forest types, Næsset et al. (2011) used lidar-based stratified estimation for aboveground biomass, albeit with model-assisted rather than SRS estimation within strata, and achieved RE > 5.

Finally, the financial aspects of the results merit attention. First, at the time of the study, the cost of acquiring lidar data was approximately \$75 (US)/km^2. Thus, for the entire study area (Section 2), the total lidar cost was approximately \$180,000 (US). Second, $RE_{VOL} = 3.0$ means that to achieve the same reduction in variance without the lidar-based stratifications, the sample size would have to be increased by factor of 3.0, i.e., the current sample size of $n = 266$ would have to be increased by $2 \cdot n = 532$ additional plots. At the time of the study, the Norwegian NFI estimated the cost of measuring a plot in the study area as approximately \$500 (US)/plot. However, many variables other than species, diameter, and height which are needed to estimate tree volume are measured on these plots. Therefore, if VOL were the only variable of interest for the additional plots, and acknowledging the fixed costs of measuring a plot such as the cost of travel to and from the plot location, the cost of measuring these additional plots may be reasonably assumed to be approximately \$350/plot. The result is that in terms of plot measurement costs, achieving the same variance reduction as the lidar-based stratifications produced would cost approximately \$180,000 (US). Thus, the costs associated with the two approaches are quite similar. In terms of future prospects, however, the costs of labor-intensive plot measurement are likely to increase, whereas lidar costs are likely to decrease. If so, lidar-based stratification becomes much more cost efficient.

Multiple factors affect the cost efficiency of using lidar-based stratifications to increase the precision of NFI estimates. First, lidar data are often acquired for purposes such as topographic mapping rather than specifically for inventory purposes. For these cases, joint funding of the acquisition reduces the inventory share of the costs and increases cost efficiency. Second, if lidar data were acquired specifically for stratification or similar forest applications, greater design efficiencies could be realized. For example, the 0.7 pulses/m^2 density of the lidar data used for this study was dictated by an authority external to the NFI. However, evidence in the literature indicates densities on the order of 0.2 pulses/m^2 would suffice for forest applications (Gobakken & Næsset, 2008). As a second example, greater increases in precision could be realized by optimizing the plot size for use with lidar data (Næsset et al., 2011). Thus, the combination of cost-sharing to reduce lidar acquisition costs and design efficiencies to increase statistical relative efficiencies contribute to overall greater cost efficiency.

5. Conclusions

Six conclusions may be drawn from the study. First, stratifications based on lidar-based predictions of growing stock volume are effective at reducing variances of estimates of mean growing stock volume. Second, extrapolation problems with the linear model made its predictions less effective as a source of information on which to base stratifications. Third, stratifications whose strata boundaries were selected to maximize either RE_{VOL} or $RE_{FOR} + RE_{VOL}$ were more effective than stratifications with equal size strata. Fourth, the stratifications reduced variances for estimates of mean growing stock volume more than variances for estimates of mean proportion forest area, often by factors of 2.0 or greater. Fifth, for this study, stratifications based on nonlinear logistic model predictions with 4–6 strata and with strata boundaries selected to maximize either RE_{VOL} or $RE_{FOR} + RE_{VOL}$ were superior. Sixth, for this study, the cost of acquiring lidar data to support stratified estimation to reduce variances for estimates of mean growing stock volume was comparable to the cost of increasing sample sizes for the same purpose. However, the combination of expected future decreases in lidar costs, cost-sharing, and greater design efficiencies suggest that lidar-based stratifications will be a cost-effective method for increasing the precision of NFI estimates of growing stock volume.

Table 2c
Estimates for stratifications based on nonlinear logistic model predictions for six strata with strata boundaries selected to maximize RE_{VOL}.

Stratum	Stratum weight	Number of plots	Proportion of plots	Proportion forest area (FOR)		Growing stock volume (VOL)	
				Mean	SE	Mean	SE
1	0.397	122	0.459	0.607	0.044	13.242	1.818
2	0.195	35	0.132	1.000	0.000	61.686	5.811
3	0.225	60	0.226	0.983	0.017	91.882	7.715
4	0.116	25	0.094	1.000	0.000	205.909	12.563
5	0.047	13	0.049	0.923	0.077	332.975	39.541
6	0.020	11	0.041	0.909	0.091	144.531	41.276
Total	1.000	266	1.000	0.835	1.470	80.425	3.188

Acknowledgments

The authors thank Mr. Liviu Theodor Ene for assistance with preparation of the remotely sensed data.

References

Agresti, A. (2007). *An introduction to categorical data analysis.* Hoboken, NJ: Wiley-Interscience.
Anderson-Sprecher, R. (1994). Model comparisons and R^2. *The American Statistician*, *48*(2), 113–117.
Anonymous (1999). *Pinnacle user's manual.* San Jose, CA: Javad Positioning Systems.
Anonymous (2005). *TerraScan user's guide.* Finland: Terrasolid Ltd. Jyvaskyla Available at: http://www.terrasolid.fi/system/files/tscan.pdf. Last date accessed: 2 October 2006.
Axelsson, P. (2000). DEM generation from laser scanner data using adaptive TIN models. *International Archives of Photogrammetry and Remote Sensing*, *33*(B4), 111–118.

Bickford, C. A. (1953). *Increasing the efficiency of airphoto forest surveys by better definition of classes.* Station Paper No. 58, Northeastern Forest Experiment Station, Forest Service, U.S. Department of Agriculture. Upper Darby, Pennsylvania 9 pp.

Bickford, C. A. (1960). A test of continuous inventory for National Forest management based upon aerial photographs, double sampling, and remeasured plots. In Anonymous (Ed.), *Proceedings of the Society of American Foresters Meeting, November 15-19, 1959, San Francisco, California* (pp. 143–148). Washington, DC: Society of American Foresters.

Braastad, H. (1966). Volume tables for birch. *Meddelelser fra det norske skogforsøksvesen*, *21*, 23–78 (In Norwegian with English summary).

Brantseg, A. (1967). Volume functions and tables for Scots pine. South Norway. *Meddelelser fra det norske skogforsøksvesen*, *22*, 689–739 (In Norwegian with English summary).

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

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

Dahlke, M., Breidt, F. J., Opsomer, J. D., & Van Keilegom, I. (2012). Nonparametric endogenous post-stratification estimation. *Statistica Sinica*. http://dx.doi.org/10.5705/ss.2011/272.

Fitje, A., & Vestjordet, E. (1977). Stand height curves and new tariff tables for Norway spruce. *Reports of the Norwegian Forest Research Institute*, *34*, 23–62 (In Norwegian with English summary.).

Frazer, G. W., Magnuissen, S., Wulder, M. A., & Niemann, K. O. (2011). Simulated impact of sample plot size and co-registration error on the accuracy and uncertainty of LiDAR-derived estimates of forest stand biomass. *Remote Sensing of Environment*, *115*, 636–649.

Gobakken, T., & Næsset, E. (2008). Assessing effects of laser point density, ground sampling intensity, and field plot sample size on biophysical stand properties derived from airborne laser scanner data. *Canadian Journal of Forest Research*, *38*, 1095–1109.

Gregoire, T. G., & Valentine, H. T. (2008). *Sampling strategies for natural resources and the environment.* Boca Raton: Chapman & Hall/CRC 474 pp.

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. 1999 November 2-3; San Antonio, Texas. General Technical Report NC-213.* (pp. 20–27) St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station Available at: http://fia.fs.fed.us/symposium/proceedings/. Last accessed: 05 March 2012.

Holden, G. R., Nelson, M. D., & McRoberts, R. E. (2005). Accuracy assessment of FIA's nationwide biomass mapping products: Results from the NC-FIA region. In R. E. McRoberts, G. A. Reams, P. C. Van Deusen, W. H. McWilliams, & C. J. Cieszewski (Eds.), *Proceedings of the fifth annual Forest Inventory and Analysis symposium. 2003 November 18-20; New Orleans, Louisiana. General Technical Report WO-69.* (pp. 139–148) Washington, DC: U.S. Department of Agriculture, Forest Service Available at: http://fia.fs.fed.us/symposium/proceedings/. Last accessed: 05 March 2012.

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *47*, 663–685.

Kendall, R. H., & Sayn-Wittgenstein, L. (1961). A test of the effectiveness of air-photo stratification. *Forestry Chronicle*, *37*(4), 350–355.

Kish, L. (1965). *Survey sampling.* New York: John Wiley & Sons.

Kraus, K., & Pfeifer, N. (1998). Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *53*, 193–203.

Lawrence, P. R., & Walker, B. B. (1954). Methods and results of forest assessment using random sampling units in photo-interpreted strata. *Australian Forestry*, *18*(2), 107–127.

Li, Y., Andersen, H. -E., & McGaughey, R. (2008). A comparison of statistical methods for estimating forest biomass from Light Detection and Ranging data. *Western Journal of Applied Forestry*, *23*(4), 223–231.

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.), *Proceedings of the joint meeting of The 6th International Symposium On Spatial Accuracy Assessment In Natural Resources and Environmental Sciences and The 15th Annual Conference of The International Environmetrics Society. June 28 – July 1 2004, Portland, Maine, USA* Available at: http://www.spatial-accuracy.org/liknes2004accuracy. Last accessed: 05 March 2012.

Loetsch, F., & Haller, K. E. (1964). *Forest Inventory: vol 1. Statistics of forest inventory and information from aerial photographs*. Munchen, Germany: BLV Verlagsgesselschaft.

Macpherson, A. F. (1962). Testing the efficiency of air-photo stratification used in the New Brunswick forest inventory. *The Forestry Chronicle*, *38*(4), 450–458.

McRoberts, R. E. (2010). The effects of rectification and Global Positioning System errors on satellite image-based estimates of forest area. *Remote Sensing of Environment*, *114*, 1710–1717.

McRoberts, R. E. (2012). Estimating forest attribute parameters for small areas using nearest neighbors techniques. *Forest Ecology and Management*, *272*, 3–12.

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., Tomppo, E. O., & Næsset, E. (2010). Advances and emerging issues in national forest inventories. *Scandinavian Journal of Forest Research*, *25*, 368–381.

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.

Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, *80*, 88–99.

Næsset, E., & Gobakken, T. (2008). Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sensing of Environment*, *112*, 3079–3090.

Næsset, E., Gobakken, T., Solberg, S., Gregoire, T. G., Nelson, R., Ståhl, G., et al. (2011). Model-assisted regional forest biomass estimation using LiDAR and InSAR as auxiliary data: A case study from a boreal forest area. *Remote Sensing of Environment*, *115*, 3599–3614.

Norwegian Meteorological Institute (2012). Climate data for municipalities. Available at: http://retro.met.no/observasjoner/hedmark/normaler_for_kommune_430.html?kommuner Last accessed: 20 June 2012.

Poso, S. (1972). A method of combining photo and field samples in forest inventory. *Communicationes Instituti Forestalis Fenniae*, *76*.

Poso, S., Hame, T., & Paananen, R. (1984). A method of estimating the stand characteristics of a forest component using satellite imagery. *Silva Fennica*, *18*, 261–292.

Poso, S., Paananen, R., & Simila, M. (1987). Forest inventory by compartments using satellite imagery. *Silva Fennica*, *21*, 69–94.

Reutebuch, S. E., McGaughey, R. J., Andersen, H. -E., & Carson, W. W. (2003). Accuracy of a high-resolution lidar terrain model under a conifer forest canopy. *Canadian Journal of Remote Sensing*, *29*, 527–535.

Särndal, C., Swensson, B., & Wretman, J. (1992). *Model-assisted survey sampling.* New York: Springer-Verlag.

Scott, C. T., Bechtold, W. A., Reams, G. A., Smith, W. D., Westfall, J. A., Hansen, M. H., et al. (2005). Sample-based estimators used by the Forest Inventory and Analysis National Information Management System. In W. A. Bechtold, & P. L. Patterson (Eds.), *The Enhanced Forest Inventory and Analysis program - national sampling design and estimation procedures. General Technical Report SRS-80*. Asheville, NC: USDA Forest Service, Southern Research Station 85 pp.

Scott, J. M., Davis, F., Csuti, B., Noss, R., Butterfield, B., Groves, C., et al. (1993). Gap analysis: a geographical approach to protection of biological diversity. *Wildlife Monograph*, *123*, 1–41.

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

Tomter, S. M., Hylen, G., & Nilsen, J. -E. (2010). Development of Norway's National Forest Inventory. In E. Tomppo, T. Gschwantner, M. Lawrence, & R. E. McRoberts (Eds.), *National Forest Inventories - Pathways for Common Reporting* (pp. 411–424). : Springer.

Vanclay, J. K., & Skovsgaard, J. P. (1997). Evaluating forest growth models. *Ecological Modelling*, *98*, 1–12.

Vestjordet, E. (1967). Functions and tables for volume of standing trees. Norway spruce. *Meddelelser fra det norske skogforsøksvesen*, *22*, 539–574 (In Norwegian with English summary).

Vestjordet, E. (1968). Merchantable volume of Norway spruce and Scots pine based on relative height and diameter at breast height or 2.5 m above stump level. *Meddelelser fra det norske skogforsøksvesen*, *25*, 411–459 (In Norwegian with English Summary.).

Vogelmann, J. E., Howard, S. M., Yang, L., Larson, C. R., Wylie, B. K., & Van Driel, N. (2001). Completion of the 1990s National Land Cover Data Set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric Engineering and Remote Sensing*, *67*, 650–662.

Westfall, J. A., Patterson, P. L., & Coulston, J. W. (2011). Post-stratified estimation: within-strata and total sample size recommendations. *Canadian Journal of Forest Research*, *41*, 1130–1139.

Zhao, K., Popescu, S., & Nelson, R. (2009). Lidar remote sensing of forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sensing of Environment*, *113*, 182–196.