Remote Sensing of Environment 115 (2011) 3165–3174

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Parametric, bootstrap, and jackknife variance estimators for the k-Nearest Neighbors technique with illustrations using forest inventory and satellite image data

Ronald E. McRoberts [a,*], Steen Magnussen [b], Erkki O. Tomppo [c], Gherardo Chirici [d]

[a] Northern Research Station, U.S. Forest Service, Saint Paul, Minnesota USA
[b] Pacific Forestry Centre, Canadian Forest Service, Vancouver, British Columbia, Canada
[c] Finnish Forest Research Institute, Vantaa, Finland
[d] University of Molise, Isernia, Italy

ARTICLE INFO

Article history:
Received 13 May 2011
Received in revised form 6 June 2011
Accepted 7 July 2011
Available online 27 August 2011

Keywords:
Model-based inference
Cluster sampling

ABSTRACT

Nearest neighbors techniques have been shown to be useful for estimating forest attributes, particularly when used with forest inventory and satellite image data. Published reports of positive results have been truly international in scope. However, for these techniques to be more useful, they must be able to contribute to scientific inference which, for sample-based methods, requires estimates of uncertainty in the form of variances or standard errors. Several parametric approaches to estimating uncertainty for nearest neighbors techniques have been proposed, but they are complex and computationally intensive. For this study, two resampling estimators, the bootstrap and the jackknife, were investigated and compared to a parametric estimator for estimating uncertainty using the k-Nearest Neighbors (k-NN) technique with forest inventory and Landsat data from Finland, Italy, and the USA. The technical objectives of the study were threefold: (1) to evaluate the assumptions underlying a parametric approach to estimating k-NN variances; (2) to assess the utility of the bootstrap and jackknife methods with respect to the quality of variance estimates, ease of implementation, and computational intensity; and (3) to investigate adaptation of resampling methods to accommodate cluster sampling. The general conclusions were that support was provided for the assumptions underlying the parametric approach, the parametric and resampling estimators produced comparable variance estimates, care must be taken to ensure that bootstrap resampling mimics the original sampling, and the bootstrap procedure is a viable approach to variance estimation for nearest neighbor techniques that use very small numbers of neighbors to calculate predictions.

Published by Elsevier Inc.

1. Introduction

Nearest neighbors techniques have emerged as a useful and popular approach for forest inventory mapping and areal estimation, particularly when used with satellite imagery as ancillary data. Nearest neighbors techniques are multivariate, non-parametric approaches to estimation based on similarity in a space of ancillary variables between a population unit for which an estimate is required and population units for which observations are available. Applications have been reported for a large number of countries in Europe, North and South America, Asia, and Africa (Fig. 1) (McRoberts et al., 2010). A bibliography of nearest neighbors papers is available at: http://blue.for.msu.edu/NAFIS/biblio.html.

Recent nearest neighbors investigations have shifted from simple descriptions of applications to more foundational work on efficiency and inference. McRoberts (2009a) reported diagnostic tools for use with univariate continuous response variables. Finley et al. (2006) and Finley and McRoberts (2008) investigated enhanced search algorithms for identifying nearest neighbors. Tomppo and Halme (2004), Tomppo et al. (2009), and McRoberts (2009b) used a genetic algorithm approach to optimize the weights for ancillary variables in the distance metric. Magnussen et al. (2010b) developed a calibration technique that improves predictions when nearest neighbors are relatively distant.

Despite these advances, the full potential of nearest neighbors techniques cannot be realized unless they can be used to construct valid statistical inferences. For probability-based inference, McRoberts et al. (2002) illustrated use of nearest neighbors techniques to support stratified estimation, and Baffetta et al. (2011, 2009) described use of nearest neighbors techniques with the model-assisted difference estimator (Särndal et al., 1992). For model-based inference, Magnussen et al. (2009) developed an estimator for mean square error, and Magnussen et al. (2010a) reported a balanced repeated replications (BRR) estimator of variance. McRoberts et al. (2007) derived a parametric nearest neighbors variance estimator for areal means from the conceptual assumptions underlying k-NN estimation. However, the latter estimator is complex, computationally intensive, and is based on assumptions that have not been closely investigated.

* Corresponding author. Tel.: +1 651 649 5174; fax: +1 651 649 5140.
E-mail address: rmcroberts@fs.fed.us (R.E. McRoberts).

0034-4257/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.rse.2011.07.002

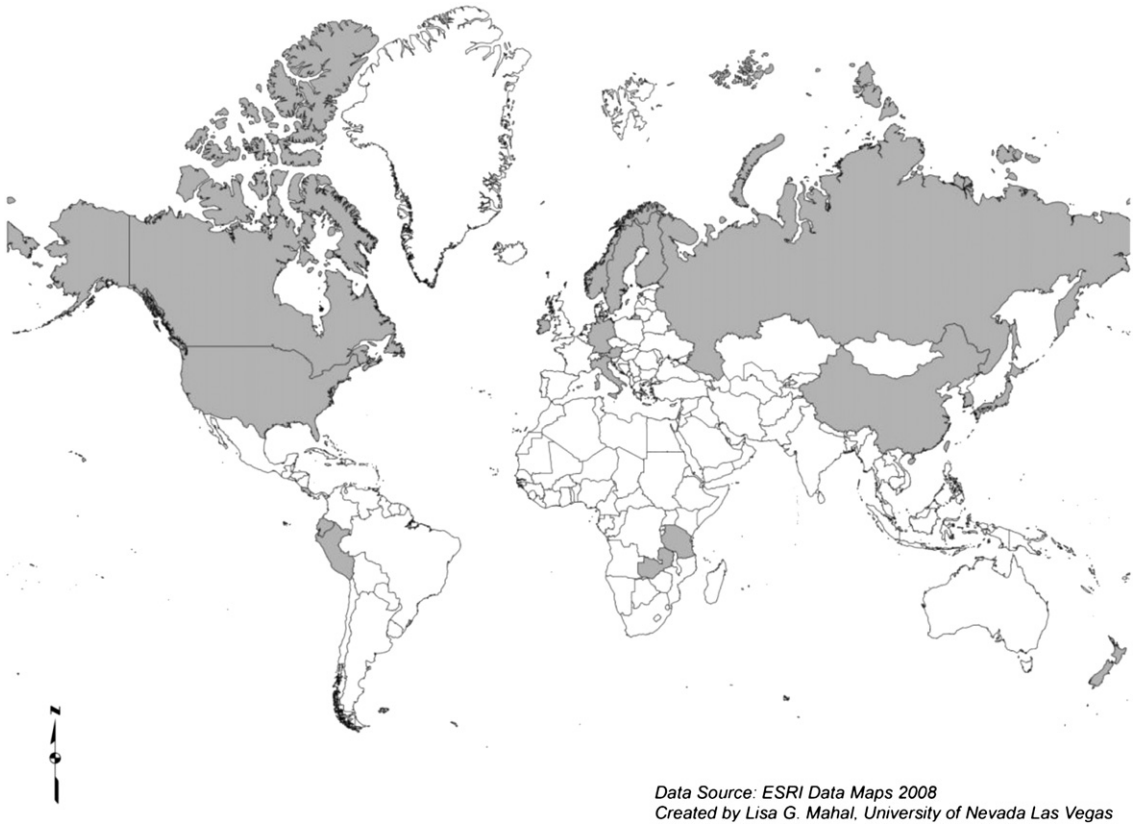

Fig. 1. Nearest neighbor applications have been reported for countries depicted in gray.

Resampling procedures are particularly well-suited for complex and non-parametric model applications and for applications requiring assumptions whose validity is difficult to assess. The jackknife resampling procedure was first proposed by Quenouille (1949) for bias reduction and by Tukey (1958) for variance estimation. The jackknife procedure produces estimates of properties of statistical estimators by sequentially deleting observations from the original sample and then re-calculating estimates using the reduced samples. The bootstrap resampling procedure was invented by Efron (1979, 1981, 1982) and further improved by Efron and Tibshirani (1994). With bootstrapping, properties of an estimator, such as its variance, are estimated via repeated sampling with replacement from an approximating distribution such as the empirical distribution of the sample observations.

For survey applications, of which forest inventory is an example, resampling methods have generated moderate interest. Shao (1996) reviewed resampling methods for sample survey applications and noted that among these methods BRR, the jackknife, and the bootstrap are the most popular. Rao (2007) briefly reviewed resampling methods for survey applications and reported use of the jackknife and bootstrap methods for small area estimation using a linear model. Chambers and Dorfman (2003) describe how a design-based bootstrap can be applied to model-based sample survey inference.

Literature on the use of resampling methods in conjunction with nearest neighbors techniques is sparse. For classification applications, Steele and Patterson (2000) proposed a weighted nearest neighbors technique based on resampling ideas, Chen and Shao (2001) reported use of jackknife methods to impute missing values for estimation of a design-based mean, and Shao and Sitter (1996) used bootstrapping in conjunction with imputation for missing data. Magnussen et al. (2010a) reported a BRR application using nearest neighbors techniques with forest inventory and satellite image data. Although the latter estimator performed well, sample sizes for which it can be applied are limited. In addition, Field and Welsh (2007, page 389) reported that for clustered data, the bootstrap estimator produces more reliable estimates of standard errors than the BRR estimator. Nothdurft et al. (2009) reported using bootstrap procedures to estimate variances of estimates of forest variables obtained using nearest neighbors techniques but did not elaborate on implementation of the resampling procedures. In summary, so few reports have been published on resampling variance estimators for use with nearest neighbors techniques that no consensus has emerged regarding their general applicability or utility. In addition, adaptation of resampling methods to accommodate cluster sampling, a feature of many forest inventory programs, has rarely been addressed.

The overall objective of the study was to compare parametric, bootstrap, and jackknife methods for estimating the variances of estimates of small area means of forest attributes obtained using the k-Nearest Neighbors (k-NN) technique. The investigations focused on three particular technical objectives: (1) to evaluate the assumptions underlying a parametric approach to estimating k-NN variances; (2) to assess the utility of the bootstrap and jackknife methods with respect to the quality of variance estimates, ease of implementation, and computational intensity; and (3) to investigate adaptation of resampling methods to accommodate cluster sampling. The investigations were based on Landsat Thematic Mapper (TM) imagery and forest inventory plot data from Finland, Italy, and the United States of America (USA).

2. Data

Four datasets, one each for Finland and Italy and two for the USA, as described below were used for the study.

2.1. North Karelia, Finland

The study area is a portion of the North Karelia forestry center in eastern Finland. Landsat 7 ETM+ data for rows 16 and 17 of path 186 were obtained for June 2000. Raw spectral data for the seven Landsat 7 ETM+ bands were used. Within the study area, an 8-km × 8-km area of

interest (AOI) was selected. Ground data were obtained for plots measured in the summer of 2000 as part of the 9th Finnish National Forest Inventory (NFI). A systematic sampling design was used, and plots were configured into clusters of 18 temporary plots or 14 permanent plots, 300 m apart, on the sides of rectangular tracts. Plot locations were obtained using geographic positioning system (GPS) receivers. Plot trees were selected using Bitterlich sampling with basal area factor 2 and a maximum plot radius of 12.52 m. Growing stock volumes of individual measured trees were estimated using statistical models, aggregated at plot-level, expressed as volume per unit area (m^3/ha), and considered as observations without error. Plot-level volumes were used for clusters for which observations were available for all plots, i.e., 257 clusters of 18 temporary plots and 19 clusters of 14 permanent plots. Plot data were associated with the spectral band values for pixels containing field plot centers. Spatial correlation among plot-level volume observations was evaluated using a variogram approach and was found to be non-negligible among observations for plots from the same cluster but negligible among observations for plots from different clusters (Sections 3.5.1, 4.1). Land use in the study area includes agriculture but is mostly forestry with Scots pine (*Pinus sylvestris* L.), Norway spruce (*Picea abies* (L.) Karst.), and birch (*Betula* spp.) being most common but with some aspen (*Populus tremula* L.) and alder (*Alnus* spp.). Additional details are available in Tomppo and Halme (2004). For future reference, the dataset is designated *North Karelia*.

2.2. Molise, Italy

The study area is in the administrative region of Molise in central Italy. Landsat ETM+ data for row 31, path 90, were obtained for July 2006, and the raw spectral data for bands 1–5, and 7 were used for analyses. Within the study area, a 7.5 km × 7.5 km AOI was selected. Field data were collected between 2005 and 2009 as part of a local forest biomass inventory using a systematic, unaligned sampling design and circular sample plots of radius 13.82 m. On average, the distance between plots was approximately 1 km. All trees with diameter at breast height (dbh) (1.30 m) of 3 cm or greater were measured. Growing stock volumes of individual measured trees were estimated using statistical models, aggregated at plot-level, expressed as volume per unit area (m^3/ha), and considered to be observations without error. Plot observations were associated with the spectral band values for pixels containing field plot centers. Spatial correlation among plot-level volume observations was evaluated using a variogram approach and was found to be negligible at distances separating plots (Sections 3.5.1, 4.1). Most forests in the Molise region are dominated by deciduous oaks (*Quercus cerris* and *Quercus pubescens*) and montane beech (*Fagus sylvatica*) which account for approximately 60% and 10% of the forest area, respectively. For future reference, the dataset is designated *Molise*.

2.3. Minnesota, USA

The study area was defined by the portion of the row 27, path 27, Landsat scene in northern Minnesota, USA. Imagery was acquired for three dates corresponding to early, peak, and late seasonal vegetative stages: April 2000, July 2001, and November 1999. Spectral data in the form of the normalized difference vegetation index (NDVI) transformation (Rouse et al., 1973) and the three tassel cap (TC) transformations (brightness, greenness, and wetness) (Crist & Cicone, 1984; Kauth & Thomas, 1976) for each of the three image dates were used. Within the study area, centers for 15 AOIs, each 8 km × 8 km, were selected using a systematic grid.

Data were obtained for plots established by the Forest Inventory and Analysis (FIA) program of the U.S. Forest Service which conducts the NFI of the USA. The program has established field plot centers in permanent locations using a sampling design that produces an equal probability sample (Bechtold & Patterson, 2005; McRoberts et al., 2005). Each FIA plot consists of four 7.32-m (24-ft) radius circular subplots that 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. Although the FIA program characterizes the individual sample units as subplots within plots, for the sake of consistency of terminology for this study they are henceforth characterized as plots within plot clusters. In general, centers of forested, partially forested, or previously forested clusters are determined using GPS receivers, whereas centers of non-forested clusters are verified using aerial imagery and digitization methods. Data were obtained between 1999 and 2003 for the 3116 plots in 779 clusters whose centers were within 15 km of the centers of the 15 AOIs. Field crews observe species and measure dbh (1.37 m, 1.5 ft) and height for all trees with dbh ≥ 12.7 cm (5 in.). Growing stock volumes of individual measured trees were estimated using statistical models, aggregated at plot-level, expressed as volume per unit area (m^3/ha), and considered to be observations without error. Plot-level volume observations were combined with the values of spectral transformations for pixels containing plot centers. Spatial correlation among volume observations was negligible for plots from different clusters but non-negligible for plots from the same cluster (Sections 3.5.1, 4.1). For this study, two FIA-based datasets were constructed: (1) the *Minnesota-central* dataset was restricted to data for the 779 central plots of clusters as a means of avoiding the necessity of dealing with spatial correlation among observations for plots from the same cluster, and (2) the *Minnesota-all* dataset consisted of data for all 3116 plots.

3. Methods

3.1. Nearest neighbors techniques

In the terminology of nearest neighbors techniques, the ancillary variables are characterized as *feature variables*, and the spaced defined by the feature variables is characterized as the *feature space*; the set of population units for which observations of both response and feature variables are available is characterized as the *reference set*; and the set of population units for which estimates are required is characterized as the *target set*. With nearest neighbors techniques, the estimate, $\tilde{y}_i$, for the ith target unit is calculated as,

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

where $\{y_j^i; j = 1,\ldots,k\}$ is the set of observations for the k reference set units nearest in feature space to the ith target set unit with respect to a selected distance metric, d. The weights, $\{w_j^i\}$, are often calculated as,

$$w_j^i = d_{ij}^{-t}, \tag{2}$$

where d_{ij} is the distance in feature space between the ith target set unit and the jth nearest reference set unit with respect to the distance metric, d, and $0 \le t \le 2$. For this study, $t = 0$ which allocates equal weights to observations for the nearest reference set observations in Eq. (1).

The distance metric, d, may often be expressed in matrix form as,

$$d_{ij} = \left[\left(\mathbf{X}_i - \mathbf{X}_j\right)' \mathbf{M} \left(\mathbf{X}_i - \mathbf{X}_j\right)\right]^{1/2}, \tag{3}$$

where $\mathbf{X}_i$ and $\mathbf{X}_j$ are the vectors of feature variables for the ith reference and jth target set units, respectively, and $\mathbf{M}$ is a square matrix. When $\mathbf{M}$ is the identity matrix, Euclidean distance results; when $\mathbf{M}$ is a non-identity diagonal matrix, weighted Euclidean distance results; and when $\mathbf{M}$ is the inverse of the covariance matrix of the feature variables, Mahalanobis distance results. Distance metrics that are optimized using

observations of the response variables include the Canonical Correlation Analysis metric (LeMay & Temesgen, 2005; Moeur & Stage, 1995), the Canonical Correspondence metric (Ohmann & Gregory, 2002), and the Fuzzy, Multiple regression, and Nonparametric metrics (Chirici et al., 2008). For this study, the unweighted Euclidean distance metric was used.

3.2. Inference

The Oxford English Dictionary definition of *infer* relevant for scientific investigations is "to accept from evidence" (Simpson & Weiner, 1989). Because evidence in the form of complete enumerations of most natural resources populations is prohibitively expensive, if not physically impossible, statistical procedures have been developed to infer values for population parameters from estimates based on observations from a sample of population units. Thus, inference requires expression of the relationship between a population parameter and its estimate in probabilistic terms (Dawid, 1983) such as in the form of familiar $1-\alpha$ confidence intervals. Statistical inference generally proceeds from either a probability-based (design-based) or a model-based framework. Probability-based inference is based on the assumption of one and only one possible value for each population unit and relies for validity on randomization in the selection of population units for the sample. However, for small areas with small sample sizes, estimates of population means often deviate substantially from true values, and variances are often unacceptably large.

Model-based approaches to inference, which are often more amenable to small area estimation, are based on the assumption that the observation for a population unit is a random variable whose value is a single realization from a distribution of possible values rather than a constant as is the case for probability-based inference. Randomization for model-based inference enters through the random realization of observations from the distribution of possible observations for each population unit rather than from the random selection of population units into the sample as is the case for probability-based inference. As a result, model-based inference is often characterized as conditional on the sample. With model-based inference, a model or prediction procedure is used to estimate the mean of the distribution for each population unit. Thus, the validity of model-based inference is based on properties of the model and its fit to the data, not the properties of the sample as is the case for probability-based inference. An important aspect of model-based inference is that when the model is correctly specified, the estimator is generally approximately unbiased (Lohr, 1999). However, when the model is misspecified, the adverse effects may be substantial (Hansen et al., 1983; Royall & Herson, 1973). Thus, model-based inferential methods often include an assessment of the quality of the fit of the model to the data. For use with satellite data, model-based estimators produce maps as by-products, population estimates that are compatible with the aggregation of mapping unit predictions, and viable estimates for small areas; however, they cannot be assumed to be unbiased, and the computational intensity may be substantial, particularly for variance estimators. For this study, model-based estimators were used because of their utility for both large and small area estimation.

3.3. Model-based inference for nearest neighbors techniques

In the context of model-based inference, μ_i and σ_i denote the mean and standard deviation, respectively, of the distribution of the response variable, Y, for the ith population unit. An observation of Y for the ith unit is therefore expressed as,

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

where ε_i is the random deviation of the observation, y_i, from its mean, μ_i. For nearest neighbors applications, the estimate of μ_i is the k-NN estimate from Eq. (1), i.e., $\hat{\mu}_i = \tilde{y}_i$. The population mean, μ, is then estimated as,

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N}\hat{\mu}_i = \frac{1}{N}\sum_{i=1}^{N}\tilde{y}_i, \tag{5}$$

where N is the population size. The variance of $\hat{\mu}$ can be estimated as,

$$\begin{aligned} \hat{Var}(\hat{\mu}) &= \hat{Var}\left(\frac{1}{N}\sum_{i=1}^{N}\hat{\mu}_i\right) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{Cov}\left(\hat{\mu}_i,\hat{\mu}_j\right) \\ &= \frac{1}{N^2}\left[\sum_{i=1}^{N}\hat{Var}(\hat{\mu}_i) + 2\sum_{i}^{N}\sum_{<j}^{N}\hat{Cov}\left(\hat{\mu}_i,\hat{\mu}_j\right)\right] \end{aligned} \tag{6}$$

as per McRoberts et al. (2007) who provide a derivation and examples.

McRoberts et al. (2007) derived k-NN estimators for the variance and covariances necessary for use with Eq. (6). The estimator for σ_i^2 is,

$$\hat{\sigma}_i^2 = \frac{\sum_{j=1}^{k}\left(y_j^i - \hat{\mu}_i\right)^2}{k - \frac{1}{k}\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\hat{\rho}_{j_1 j_2}}, \tag{7a}$$

where $\rho_{j_1 j_2}$ is the spatial correlation between observations of Y for the j_1th and j_2th neighbors of the ith target set unit. In the absence of spatial correlation, Eq. (7a) reduces to the more familiar,

$$\hat{\sigma}_i^2 = \frac{1}{k}\sum_{j=1}^{k}\left(y_j^i - \hat{\mu}_i\right)^2. \tag{7b}$$

An estimator for $Cov\left(\hat{\mu}_i,\hat{\mu}_j\right)$ is,

$$\hat{Cov}\left(\hat{\mu}_i,\hat{\mu}_j\right) \approx \frac{\hat{\sigma}_i\hat{\sigma}_j}{k^2}\sum_{l_i=1}^{k}\sum_{l_j=1}^{k}\hat{\rho}_{l_i l_j}, \tag{8a}$$

where l_i and l_j index the neighbors nearest in the reference set to the ith and jth target set elements. In the absence of spatial correlation Eq. (8a) reduces to,

$$\hat{Cov}\left(\hat{\mu}_i,\hat{\mu}_j\right) \approx m_{ij}\frac{\hat{\sigma}_i\hat{\sigma}_j}{k^2}, \tag{8b}$$

where m_{ij} is the number of common nearest neighbors used to calculate $\hat{\mu}_i$ and $\hat{\mu}_j$, and,

$$\hat{Var}(\hat{\mu}_i) = \hat{Cov}(\hat{\mu}_i,\hat{\mu}_i) \approx m_{ii}\frac{\hat{\sigma}_i^2}{k^2} = \frac{\hat{\sigma}_i^2}{k}, \tag{9}$$

with the last step being the result of $m_{ii} = k$. Estimates of σ_i^2 may be obtained by substituting into Eqs. (7a) and (7b) estimates, $\hat{\mu}_i = \tilde{y}_i$, from Eq. (1) and estimates, $\hat{\rho}_{j_1 j_2}$, may be obtained using a variogram approach as outlined by McRoberts et al. (2007). Similarly, estimates $\hat{Cov}\left(\hat{\mu}_i,\hat{\mu}_j\right)$ and $\hat{Var}(\hat{\mu}_i)$ may be obtained by substituting estimates, $\hat{\sigma}_i^2$, and $\hat{\rho}_{j_1 j_2}$, into Eqs. (8a), (8b), and (9), and the estimate, $\hat{Var}(\hat{\mu})$, may be obtained by substituting $\hat{Var}(\hat{\mu}_i)$ and $\hat{Cov}\left(\hat{\mu}_i,\hat{\mu}_j\right)$ into Eq. (6). The derivations of these variance estimators by McRoberts et al. (2007, Appendices) were based on simplifying assumptions that have not been closely investigated.

3.4. Resampling methods

Estimation of $\hat{Var}(\hat{\mu})$ using resampling methods such as the bootstrap and jackknife procedures are alternatives to the parametric approach, Eq. (6), that do not depend on the assumptions underlying Eqs. (7a)–(9) other than Eq. (4).

3.4.1. The jackknife procedure

Efron and Tibshirani (1994) refer to the jackknife, proposed by Quenouille (1949), as "the original computer-based method for estimating biases and standard errors." For a dataset of size n, the jth *jackknife sample* is defined to be the original dataset with the jth data point removed. The estimate, $\hat{\mu}_{jack}^{-j}$, is obtained from the jth jackknife sample, and the jackknife population estimate is,

$$\hat{\mu}_{jack} = \frac{1}{n}\sum_{j=1}^{n}\hat{\mu}_{jack}^{-j}. \quad (10)$$

The jackknife estimate of bias is calculated as,

$$\hat{Bias}_{jack}\left(\hat{\mu}\right) = (n-1)\left(\hat{\mu}_{jack}-\hat{\mu}\right), \quad (11)$$

where $\hat{\mu}$ is the estimate obtained using the entire dataset. The corresponding variance estimate is calculated as,

$$\hat{Var}_{jack}\left(\hat{\mu}\right) = \frac{n-1}{n}\sum_{j=1}^{n}\left(\hat{\mu}_{jack}^{-j}-\hat{\mu}_{jack}\right)^2. \quad (12)$$

The multiplicative term, $n-1$, in Eqs. (11) and (12) adjusts for the similarity of the jackknife samples. An assumption underlying the jackknife procedure is that the statistic of interest is smooth in the sense that an incremental change in the data produces only an incremental change in the estimate of the statistic. Efron and Tibshirani (1994, p. 148) note that "the jackknife can fail miserably" if the statistic is not smooth. For nearest neighbors techniques, the smoothness criterion is not satisfied because a small change in the value of a feature variable could alter a set of nearest neighbors which, in turn, could produce a much different estimate for a population unit. Nevertheless, the estimate of the population mean may be sufficiently smooth to consider use of the jackknife estimator of variance.

3.4.2. The bootstrap procedure

The bootstrap resampling procedure was invented by Efron (1979, 1981, 1982) and further developed by Efron and Tibshirani (1994). An advantage of the bootstrap procedure over the jackknife procedure is that smoothness is not required. All bootstrap methods depend on the notion of a *bootstrap sample*. For modeling problems, Efron and Tibshirani (1994) describe two general approaches to constructing a bootstrap sample. First, for a sample of n pairs $(y_i, \mathbf{X}_i)$ and the corresponding empirical distribution of the pairs, each with probability 1/n, a bootstrap sample is defined to be a sample of size n drawn with replacement from the empirical distribution. This approach is characterized as *bootstrapping pairs*. However, the pairs can also be expressed as $(\hat{y}_i+\varepsilon_i, \mathbf{X}_i)$ where $\varepsilon_i = y_i-\hat{y}_i$ which leads to a second approach to bootstrapping that focuses on resampling the residuals and is characterized as *bootstrapping residuals*. With this approach, a random sample of residuals is drawn with replacement, either from the empirical distribution of residuals or a parametric model of the distribution, and the residuals are added back to the model predictions to form a bootstrap sample. An assumption underlying bootstrapping residuals is that the residuals are independently and identically distributed (iid). Therefore, in the case of heteroscedasticity, the residuals must initially be studentized using methods such as those described by McRoberts (2009a).

Regardless of how the bootstrap sample is constructed, the estimate $\hat{\mu}_{boot}^{b}$ is obtained from the b^{th} bootstrap sample, and the bootstrap population estimate is,

$$\hat{\mu}_{boot} = \frac{1}{n_{boot}}\sum_{b=1}^{n_{boot}}\hat{\mu}_{boot}^{b}, \quad (13)$$

where n_{boot} is the number of bootstrap samples. The bootstrap estimate of bias is calculated as,

$$\hat{Bias}_{boot}(\hat{\mu}) = \hat{\mu}_{boot}-\hat{\mu}, \quad (14)$$

where $\hat{\mu}$ is the estimate obtained from the original sample, and the bootstrap estimate of variance is calculated as,

$$\hat{Var}_{boot}\left(\hat{\mu}\right) = \frac{1}{n_{boot}-1}\sum_{b=1}^{n_{boot}}\left(\hat{\mu}_{boot}^{b}-\hat{\mu}_{boot}\right)^2. \quad (15)$$

The estimate of the standard error, obtained as the square root of the variance estimate from Eq. (15), is characterized as the *ideal bootstrap* estimate (Efron & Tibshirani, 1994, page 46).

Because randomization for model-based inference occurs in the realization of observations from the distribution for each population unit, bootstrapping residuals is the more intuitive approach, although for k-NN applications the iid assumption may be difficult to satisfy. For example, for population units for which k-NN estimates are zero, all residuals will be non-negative because all plot observations are non-negative, a condition that will not occur for non-zero k-NN estimates (McRoberts, 2009a, Fig. 5). Further, Efron and Tibshirani (1994, page 113) note that bootstrapping residuals is more sensitive to assumptions than bootstrapping pairs for which the only assumption is that the original pairs represent a random sample from an appropriate distribution. Fortunately for k-NN applications, Efron and Tibshirani (1994) also note that results obtained when bootstrapping pairs approach those obtained when bootstrapping residuals as sample sizes increase. McRoberts (2010) confirmed this assertion using a dataset similar to the Minnesota-central dataset used for this study, albeit with a different approach to estimation. Therefore, for this study, only bootstrapping pairs was used.

A crucial aspect of bootstrap methods is that the resampling must be from an appropriate distribution, i.e., the resampling must adequately mimic the original sampling. For the Molise and Minnesota-central datasets, the sampling designs have systematic components but may be considered simple random sampling for purposes of variance estimation and bootstrap resampling. Thus, the ideal bootstrap approach may be used for these datasets. For the North Karelia and Minnesota-all datasets, lack of independence in the selection of locations of plots from the same cluster must be mimicked in the bootstrap resampling. Field and Welsh (2007) review approaches to bootstrapping with such clustered data, provide multiple references, and develop supporting theory. Two approaches to cluster sampling are common: (1) single-stage cluster sampling, which consists of first randomly selecting clusters and then selecting all plots within clusters, and (2) two-stage cluster sampling, which consists of first randomly selecting clusters and then randomly selecting plots within clusters. For use with data obtained using single-stage cluster sampling, Field and Welsh (2007) define *cluster bootstrapping* as consisting of randomly selecting clusters with replacement, and then selecting all the plots within the selected clusters. For use with data obtained using two-stage cluster sampling, Field and Welsh (2007) define *two-stage bootstrapping* as consisting of first randomly selecting clusters with replacement and then randomly selecting plots within selected clusters with replacement.

Neither single-stage nor two-stage cluster sampling exactly characterizes the clustering features of most forest inventory sampling designs. First, plots within clusters are typically circular which effectively prohibits a tessellation of the area surrounding the cluster center into a finite number of mutually exclusive plots that completely covers the cluster area. However, for this study, clusters were considered to consist of TM pixels which do form an appropriate tessellation of the cluster area. Second, given this tesselation, two-stage cluster sampling most closely characterizes inventory cluster sampling, although plots within inventory clusters typically are not randomly

selected but rather are systematically selected, usually without any kind of randomization.

3.5. Analyses

All analyses were based on four underlying assumptions: (1) a finite population consisting of N units in the form of TM pixels, (2) feature variable data in the form of the spectral data from TM bands, or transformations of them, for all population units, (3) a sample of n population units, and (4) adequate representation of entire TM pixels by observations of sample plots whose centers are in the pixels. In the following sections, the terms *population unit* and *pixel* are used interchangeably.

For each dataset, the combination of feature variables and value of k that minimized root mean square error (RMSE) was first selected. For this combination of feature variables, and for each value of k, a simple linear regression model was fit to reference set observations as the dependent variable and the k-NN predictions as the independent variable. For the combination of feature variables that mimized RMSE, the value of k for which the estimates of the intercept and slope for the regression model were jointly closest to (0,1) was selected. This procedure simultaneously contributes to minimizing RMSE, enhancing the quality of fit of the k-NN model, and minimizing the bias of estimators using the k-NN population unit estimates.

3.5.1. Parametric variance estimation

Spatial correlation was quantified for all datasets using the variogram approach outlined by McRoberts et al. (2007) under the assumptions of stationarity, meaning that spatial correlation does not change within the study area, and isotropy, meaning that spatial correlation is the same in all directions.

Because of the double sums in the parametric estimator, Eq. (6), considerable computational intensity may be involved when calculating variance estimates. However, Eq. (6) represents a simple two-dimensional average over all N population units. As a means of reducing the computational intensity, two-dimensional grids of widths two and four pixels were superimposed on the AOI, and only the pixels with centers closest to the grid intersections were used to calculate variance estimates. When using the grids of two- and four-pixel widths, the population size, N, in Eq. (6) was replaced with the number of grid intersections. The estimates of variances obtained using the grids were compared to estimates obtained using all pixels which is equivalent to a grid with width of one pixel. Means and standard errors (SE) for the AOIs using all three grid widths were estimated using the parametric variance estimators for all four datasets and the respective AOIs.

3.5.2. Jackknife variance estimation

The jackknife resampling procedures were applied to estimate means, biases, and SEs for the Molise AOI and for the Minnesota AOIs using the Minnesota-central dataset as the k-NN reference set. The jackknife estimator was not used with the North Karelia or Minnesota-all datasets because methods to accommodate the clustering sampling feature are complex and because the jackknife entails much greater computational intensity for these larger reference sets.

3.5.3. Bootstrap variance estimation

The bootstrap resampling procedures were applied using all four datasets as reference sets, and estimates of the means, biases, and SEs were calculated. The recommendation of Efron and Tibshirani (1994) to draw at least $n_{boot} = 200$ bootstrap samples was evaluated by using $n_{boot} = 1000$ to assess the stability of $\hat{\mu}_{boot}$ and $SE_{boot}(\hat{\mu})$ with respect to the number of bootstrap samples.

For the Molise and Minnesota-center datasets, the ideal bootstrap procedure was used whereby the bootstrap samples were constructed by simply randomly selecting plots with replacement.

Because of the cluster features of the North Karelia and Minnesota-all datasets, the ideal bootstrap procedure was not expected to produce acceptable estimates. Bootstrapping for these datasets was implemented using three approaches. The first approach was cluster bootstrapping (Field & Welsh, 2007) which is intended to mimic single-stage cluster sampling and consists of randomly selecting clusters with replacement, and then selecting all the plots within the selected clusters. For the Minnesota-all dataset, clusters were selected with replacement and then all four plots within clusters were selected. For the North Karelia dataset, clusters consisted of either 14 or 18 plots. For bootstrap resampling, the numbers of 14-plot and 18-plot clusters represented in the original dataset were maintained in the bootstrap samples as a means of ensuring that the bootstrap sample size would be the same as the original sample size. Therefore, the appropriate numbers of 14-plot and 18-plot clusters were first randomly selected with replacement, and then all 14 or all 18 plots within clusters were selected. The second approach is two-stage bootstrapping (Field & Welsh, 2007) which is intended to mimic two-stage cluster sampling and consists of first randomly selecting clusters with replacement and then randomly selecting plots within selected clusters with replacement. For the Minnesota-all dataset, clusters were randomly selected with replacement and then four plots within clusters were randomly selected with replacement. For the North Karelia dataset, the appropriate numbers of 14-plot and 18-plot clusters were randomly selected with replacement, and then either 14 or 18 plots were randomly selected with replacement within the selected clusters. Finally, simply for comparison purposes, the ideal bootstrap approach was used. With this approach, the clustering aspects of the sampling designs were ignored, and plots were randomly selected with replacement without regard to their cluster associations.

3.5.4. Evaluating assumptions

The validity of the assumptions underlying Eqs. (7a), (7b), (8a) and (8b) was evaluated by comparing SEs obtained using the parametric estimator to SEs obtained using the bootstrap and jackknife estimators. Good agreement among the parametric estimates, which depend on the underlying assumptions, and the resampling estimates supports claims for the validity of the assumptions. Additional support accrues if agreement is obtained for different forest conditions for sites separated by large geographic distances.

The bootstrap approach was also used to estimate σ_i^2 for each population unit using 10, 25, 50, 100, 250, and 500 resamples. The rationale for these analyses was to further evaluate the validity of the assumptions underlying the parametric variance estimator; in particular, to assess if the estimate, $\hat{\sigma}_i^2$, obtained from Eq. (7a) which is calculated as the variance of observations for nearest neighbor population units, adequately represent variances of the distribution for the ith target population unit. For these analyses, the bootstrap variance estimate for each population unit represents the variance of the mean of the population unit distribution, not the variance of individual observations from the distribution as is the case for Eq. (7a). However, because the variance of the mean is just the variance of the observations divided by the sample size, the variances of bootstrap estimates of μ_i for population units were multiplied by the value of k to obtain the estimate of σ_i^2. The estimates of σ_i^2 obtained in this manner were used in Eqs. (8a), (8b), and (9) and in Eq. (6) to estimate $Var(\hat{\mu})$. Using the Minnesota-central dataset as the reference set, estimates of SE for the Minnesota AOIs based on estimates of σ_i^2 obtained from Eq. (7a) were compared to estimates of SE based on estimates of σ_i^2 obtained via bootstrapping. Good agreement between the sets of SE estimates provides additional support for the claim of the validity of the assumptions underlying the parametric variance estimator.

Table 1
k-NN implementation parameters for estimating mean volume per unit area (m^3/ha).

Dataset	n[a]	Number of feature variables	Minimizing RMSE				Maximizing quality of fit			
			k	b_0	b_1	RMSE	k	b_0	b_1	RMSE
Molise	181	2 of 6	19	−1.54	1.05	46.93	16	0.02	1.02	47.68
North Karelia	5900	3 of 9	45	0.00	1.00	63.77	45	0.00	1.00	63.77
Minnesota-central	779	4 of 12	28	−0.98	0.97	64.97	27	0.06	0.95	65.16
Minnesota-all	3116	5 of 12	25	0.98	0.97	65.19	34	−0.13	0.99	65.46

[a] n = sample (reference set) size.

4. Results and discussion

4.1. Implementation of the k-NN technique

For each dataset, the number of feature variables that produced the smallest RMSE was less than the total number of feature variables available (Table 1). Further, for the selected combination of feature variables, the value of k that produced estimates of slopes and intercepts for the simple linear regressions of observations versus predictions that were jointly closest to (0,1) was often different than the value of k that minimized RMSE. However, differences in RMSEs for these two values of k were small as a result of the relative flatness of the RMSE versus k curve in the vicinity of the value of k that minimizes RMSE; the latter phenomenon is characteristic of many k-NN applications.

The variogram analyses for the four datasets indicated negligible spatial correlation among observations for the Molise and Minnesota-central datasets. Therefore, Eqs. (7b) and (8b) could be used to estimate σ_i^2 and $\hat{Cov}\left(\hat{\mu}_i, \hat{\mu}_j\right)$. However, for the clustered North Karelia and Minnesota-all datasets, spatial correlation among volume observations for plots from the same cluster was non-negligible but generally did not extend to plots from different clusters. When constructing variograms, caution must be exercised because observations are often sparse for small distances for which spatial correlation is greatest.

4.2. Parametric variance estimator

Restricting the calculation of variance estimates using the parametric approach to subsets of pixels located at the intersections of 2×2 and 4×4 grids produced estimates that were nearly indistinguishable from estimates calculated using all N pixels, i.e., a 1×1 grid. This result held regardless of whether spatial correlation among observations was negligible or non-negligible (Tables 2–3). This result facilitates a substantial reduction in the computational intensity associated with use of the parametric estimator without adverse effects and broadens the appeal of the estimator. For the following discussion, references to parametric variance and SE estimates always pertain to those calculated using all N pixels, i.e., the 1×1 grid.

Of note, estimates of SEs obtained using reference sets for which spatial correlation among observations was non-negligible were somewhat sensitive to the range of correlation estimated from variograms.

Table 2
Estimates for non-clustered reference sets.

Area of interest[a]	Mean	Parametric SE (grid)			Bootstrap		Jackknife	
		1×1	2×2	4×4	Mean	SE	Mean	SE
Molise	68.75	4.87	4.86	4.84	67.57	4.91	68.72	5.43
Minnesota 1	61.35	3.34	3.34	3.33	61.25	3.37	61.35	3.36
Minnesota 2	53.54	2.87	2.87	2.88	53.57	3.02	53.54	2.88
Minnesota 3	52.96	2.97	2.96	2.96	52.78	3.02	52.96	2.95
Minnesota 4	64.37	4.37	4.37	4.40	64.16	4.17	64.37	4.37
Minnesota 5	76.44	4.40	4.40	4.42	76.09	4.39	76.43	4.44
Minnesota 6	33.30	2.09	2.09	2.10	33.46	2.18	33.33	2.19
Minnesota 7	58.99	3.08	3.08	3.10	58.98	3.20	58.99	3.16
Minnesota 8	45.85	2.31	2.31	2.31	45.98	2.44	45.85	2.34
Minnesota 9	49.70	2.91	2.89	2.90	49.81	3.01	49.70	3.02
Minnesota 10	51.69	3.92	3.92	3.89	51.60	3.92	51.69	4.05
Minnesota 11	44.98	2.85	2.85	2.85	44.99	2.90	44.99	2.91
Minnesota 12	63.18	3.65	3.66	3.65	63.22	3.84	63.18	3.77
Minnesota 13	38.11	3.06	3.06	3.05	38.31	3.15	38.11	3.16
Minnesota 14	77.17	4.37	4.37	4.39	76.66	4.36	77.16	4.46
Minnesota 15	52.40	3.17	3.18	3.16	52.59	3.26	52.40	3.24

[a] Feature variables and values of k reported in Table 1.

4.3. Comparing parametric and resampling estimators

Efron and Tibshirani (1994) recommend at least 200 bootstrap resamples to estimate means but make no recommendation regarding estimating variances or SEs. The 1000 bootstrap samples used for this study were sufficient for estimates of both means and SEs to stabilize, although the 200 bootstrap samples recommended for estimating means were not always sufficient (Fig. 2). This result should not be generalized, but rather the number of bootstrap samples necessary should be investigated separately for each application.

4.3.1. Non-clustered reference sets

The parametric, jackknife, and bootstrap estimates of both means and SEs obtained using the Molise and Minnesota-central datasets as reference sets were often similar (Table 2). For the Molise dataset, the jackknife estimate of bias was less than the bootstrap estimate, although as proportions of the parametric estimates of means, estimates of bias for both resampling approaches were less in absolute value than 0.02. As a proportion of the parametric estimate of the mean, the deviation of the bootstrap estimate of the SE from the parametric estimate was less than 0.01, whereas the deviation of the jackknife estimate was greater than 0.11. For the Minnesota AOIs, estimates of bias for the jackknife estimator were slightly smaller than for the bootstrap estimator. As proportions of the parametric estimates of means, the mean bias estimate for the bootstrap estimator was 0.0001, the mean absolute bias estimate was 0.0028, and the root mean square deviation (RMSD) was 0.0034, all of which

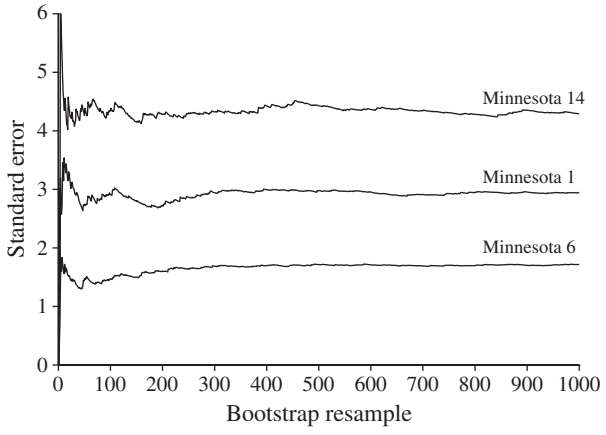

Fig. 2. Standard errors versus number of bootstrap samples for selected areas of interest representing the range of magnitudes of standard errors.

Table 3
Estimates for clustered datasets.

Area of interest[a]	Mean	Parametric SE (grid)			Bootstrap					
					Ideal		Single-stage		Two-stage	
		1×1	2×2	4×4	Mean	SE	Mean	SE	Mean	SE
North Karelia	85.88	2.02	2.02	2.02	82.47	1.48	86.27	1.70	86.17	2.03
Minnesota 1	64.30	2.71	2.72	2.71	64.43	1.85	64.35	2.68	64.29	2.95
Minnesota 2	55.47	2.15	2.15	2.15	55.63	1.42	55.65	2.07	55.63	2.18
Minnesota 3	53.38	2.36	2.36	2.36	53.54	1.60	53.36	2.23	53.32	2.47
Minnesota 4	71.43	4.20	4.20	4.22	71.76	2.83	71.27	3.81	71.36	4.38
Minnesota 5	75.56	3.97	3.97	3.97	75.73	2.75	76.14	3.83	75.96	4.03
Minnesota 6	34.82	1.69	1.68	1.66	34.93	1.17	35.29	1.55	35.38	1.72
Minnesota 7	58.92	2.31	2.32	2.33	59.02	1.58	58.99	2.27	59.03	2.39
Minnesota 8	45.48	1.76	1.76	1.77	45.57	1.23	45.68	1.79	45.69	1.85
Minnesota 9	43.97	2.02	2.01	2.03	44.01	1.42	44.25	2.08	45.69	2.18
Minnesota 10	41.85	2.43	2.43	2.40	41.84	1.65	42.17	2.50	42.20	2.81
Minnesota 11	34.80	2.80	2.79	2.79	34.81	2.00	35.57	2.43	35.66	2.90
Minnesota 12	56.91	3.02	3.02	3.01	57.00	2.09	58.12	3.23	58.18	3.37
Minnesota 13	33.66	2.03	2.04	2.02	33.60	1.39	33.87	2.06	33.98	2.14
Minnesota 14	78.26	4.27	4.28	4.29	78.26	3.01	78.46	4.00	78.23	4.29
Minnesota 15	50.87	2.33	2.33	2.35	50.90	1.53	51.36	2.32	51.44	2.43

[a] Feature variables and values of k reported in Table 1.

would be considered negligible for most applications. Values for the jackknife estimates were even smaller. However, these estimates of bias pertain only to sampling issues, not to the fitness of the k-NN procedure. Estimates of bootstrap SEs were similar to the parametric estimates; as proportions of the parametric estimates, the mean deviation was −0.0219, the mean absolute deviation was 0.0286, and the RMSD was 0.0264. Values for the jackknife estimator were slightly smaller.

4.3.2. Clustered reference sets

The parametric and bootstrap estimates of means obtained using the clustered North Karelia and Minnesota-all datasets as reference sets were similar (Table 3). For the North Karelia dataset, as proportions of the parametric mean, the single-stage and two-stage bootstrap bias estimates were less in absolute value than 0.0045, whereas the ideal bootstrap estimate of bias was approximately −0.04. As proportions of the parametric estimates of the means, the mean bootstrap estimates of bias for the Minnesota AOIs were −0.0015 for the ideal bootstrap, −0.0069 for the single-stage bootstrap estimator and −0.0096 for the two-stage estimator; the mean absolute bias estimates were 0.0018 for the ideal estimator, 0.0073 for the single-stage estimator, and 0.0099 for the two-stage estimator; and RMSDs for the bias estimates were 0.0016 for the ideal estimator, 0.0070 for the single-stage estimator, and 0.0112 for the two-stage estimator.

The results of comparing bootstrap estimates of SEs to parametric estimates depended on the approach to constructing the bootstrap samples (Table 3, Fig. 3). For both the North Karelia and the Minnesota-all datasets, the SE estimates based on the ideal approach were consistently smaller by substantial amounts than the parametric estimates and are not further discussed. For the North Karelia dataset, the deviation of the two-stage bootstrap estimate from the parametric estimate, as a proportion of the parametric estimate, was less than 0.005, whereas the deviation of the single-stage estimate was nearly 0.16. For the Minnesota-all dataset, the single-stage bootstrap SE estimates were generally smaller than the parametric estimates with mean deviation of 0.0248, mean absolute deviation of 0.0461, and RMSD of 0.0522. The two-stage bootstrap SE estimates were generally larger than the parametric estimates with mean deviation of −0.0533, mean absolute deviation of 0.0533, and RMSD of 0.0400.

Multiple causes may be proposed to explain the deviations between the parametric and two-stage bootstrap estimates of SEs for the Minnesota AOIs. First, although the mean deviations and RMSDs were small, the two-stage cluster sampling assumption does not adequately characterize the fixed geometric configuration of the FIA plot clusters. In particular, plots within clusters are selected systematically, whereas variance estimation is based on an assumption of simple random sampling of plots within clusters. Use of the simple random sampling variance estimator with systematically distributed data is known to produce overestimates of variances (Särndal et al., 1992, page 83). Second, the parametric estimates require estimation of spatial correlation using a variogram approach for which only limited data at small distances are available; thus, the variogram may have considerable uncertainty associated with it. In addition, variance estimates obtained using the parametric estimator were sensitive to estimates of the variogram parameters. Third, the stationarity and isotropy assumptions underlying estimation of spatial correlation are likely approximately satisfied over the entire study area, but the degree to which they are satisfied for smaller individual AOIs at different locations within the study area may vary considerably. These three possible causes, singly or in combination, could explain the greater deviations between the parametric and two-stage bootstrap estimates of SE. McCullagh (2000) also reported inconsistency with the two-stage bootstrap unless the number of clusters and the number of plots within clusters were large. Of particular note, for the Minnesota-all reference set, the number of plots within clusters was only four, whereas for the North Karelia reference set for which the two-stage bootstrap estimator produced

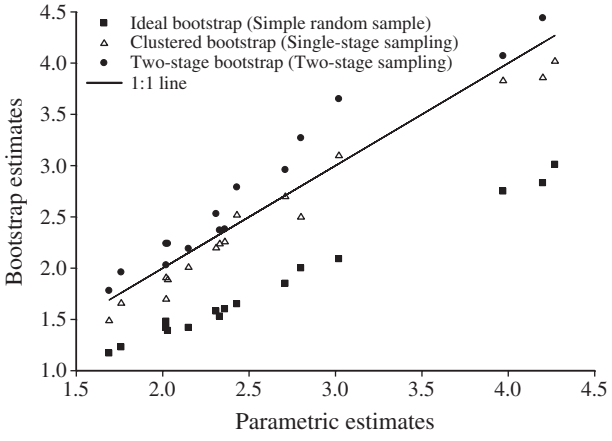

Fig. 3. Bootstrap standard error estimates versus parametric standard error estimates for Minnesota AOIs.

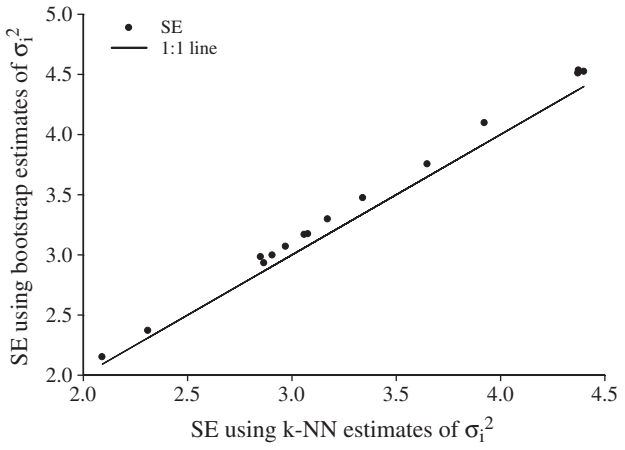

Fig. 4. Parametric estimates of standard errors (SE) for Minnesota AOIs using Minnesota-central reference set.

excellent results, the number of plots within clusters was either 14 or 18. The tendency of the two-stage bootstrap to overestimate SEs means that confidence intervals would be slightly conservative.

4.4. Evaluating assumptions

The validity of the assumptions underlying the parametric variance estimator expressed by Eqs. (7a), (7b),(8a), (8b) and (9) is supported by two results of the study: (1) the closeness of the parametric and bootstrap estimates of SEs for reference sets with both non-clustered and clustered data (Tables 2 and 3), and (2) the small changes in parametric estimates of SEs resulting from use of bootstrap rather than k-NN estimates of σ_i^2 (Fig. 4). Although SEs based on the bootstrap estimates of σ_i^2 are consistently larger, the mean absolute deviation as a proportion of the SEs based on the k-NN estimates of σ_i^2 is less than 0.035.

5. Conclusions

Four primary conclusions may be drawn from the study. First, the computational intensity associated with calculation of parametric variance estimates can be substantially reduced by using only population units located at the intersections of grids which greatly enhances the appeal of the parametric estimator. Second, the similarity in estimates of SEs obtained using the parametric, jackknife, and bootstrap estimators for different forest conditions in widely separated geographic regions lends support to a claim of validity for the assumptions underlying the parametric estimator. Additional support accrues from the finding of small differences among parametric SE estimates when σ_i^2 is estimated using the parametric approach of Eqs. (7a, 7b) or the bootstrap approach. Third, the bootstrap estimator produced good results, although issues related to cluster sampling require additional investigation. In addition, the favorable results for the bootstrap estimator constitute an argument against use of the jackknife estimator for which the underlying smoothness assumption is not satisfied for nearest neighbors estimation. Also, the number of jackknife iterations may be considerably greater than the number of bootstrap resamples necessary for large reference sets. Fourth, for nearest neighbors approaches that use small values of k, particularly $k = 1$ (e.g., LeMay & Temesgen, 2005; Ohmann & Gregory, 2002), credible estimates of σ_i^2 would be difficult if not impossible to obtain using Eqs. (7a, 7b). This difficulty would, in turn, make the parametric approach to variance estimation infeasible. However, the bootstrap approach is a viable alternative for variance estimation for these nearest neighbors approaches.

References

Baffetta, F., Corona, P., & Fattorini, L. (2011). Design-based diagnostics for k-NN estimators of forest resources. *Canadian Journal of Forest Research*, *40*(1), 59–72.

Baffetta, F., Fattorini, L., Franeschi, 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*, *113*(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. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. 85p.

Chambers, R.L., & Dorfman, A.H. (2003). Robust sample survey inference via bootstrapping and bias correction: The case of the ratio estimator. *Southampton, UK, Southampton Statistical Sciences Research Institute*, 21 pp. (S3RI Methodology Working Papers, M03/13) http://eprints.soton.ac.uk/8163/. Available at: http://eprints.soton.ac.uk/8163/01/s3ri-workingpaper-m03-13.pdf. (last accessed: January 2011).

Chen, J., & Shao, J. (2001). Jackknife variance estimation for nearest neighbour imputation. *Journal of the American Statistical Association*, *96*(453), 260–269.

Chirici, G., Barbati, A., Corona, P., Marchetti, M., Travaglini, D., Maselli, F., & Bertini, R. (2008). Non-parametric and parametric methods using satellite imagery for estimating growing stock volume in alpine and Mediterranean forest ecosystems. *Remote Sensing of Environment*, *112*, 2686–2700.

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.

Dawid, A. P. (1983). Statistical inference I. In S. Kotz, & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences*, *Vol. 4.* (pp. 89–105)New York: Wiley.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, *7*(1), 1–26.

Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, *68*, 589–599.

Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans. 38. *Society of Industrial and Applied Mathematics CBMS-NSF Monographs* 92 p.

Efron, B., & Tibshirani, R. (1994). An introduction to the bootstrap. Boca Raton, FL: Chapman and Hall/CRC 436 p.

Field, C. A., & Welsh, A. H. (2007). Bootstrapping clustered data. *Journal of the Royal Statistical Society, B*, *68*(Part 3), 369–390.

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

Finley, A. O., McRoberts, R. E., & Ek, A. R. (2006). Applying an efficient k-Nearest Neighbor search to forest attribute imputation. *Forest Science*, *52*(2), 130135.

Hansen, M. H., Madow, W. G., & Tepping, B. J. (1983). An evaluation of model-depending and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association*, *78*, 776–793.

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.

LeMay, V., & Temesgen, H. (2005). Comparison of nearest neighbor methods for estimating basal area and stems per hectare using aerial auxiliary variables. *Forest Science*, *51*, 109–119.

Lohr, S. (1999). Sampling: Design and analysis. Pacific Grove, CA: Duxbury 494 p.

Magnussen, S., McRoberts, R. E., & Tomppo, E. O. (2009). Model-based mean square error estimators for k-nearest neighbour predictions and applications using remotely sensed data for forest inventories. *Remote Sensing of Environment*, *113*, 476–488.

Magnussen, S., McRoberts, R. E., & Tomppo, E. O. (2010). A resampling variance estimator for the *k* nearest neighbours technique. *Canadian Journal of Forest Research*, *40*, 648–658.

Magnussen, S., Tomppo, E., & McRoberts, R. E. (2010). A model-assisted kNN approach to remove extrapolation bias. *Scandinavian Journal of Forest Research*, *25*, 174–184.

McCullagh, P. (2000). Re-sampling and exchangeable arrays. *Bernoulli*, *6*, 285–301.

McRoberts, R. E. (2009). Diagnostic tools for nearest neighbors techniques when used with satellite imagery. *Remote Sensing of Environment*, *113*, 489–499.

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. (2010). Probability- and model-based approaches to inference for proportion forest using satellite imagery as ancillary data. *Remote Sensing of Environment*, *114*, 1017–1025.

McRoberts, R. E., Nelson, M. D., & Wendt, D. G. (2002). Stratified estimation of forest area using satellite imagery, inventory data, and the k-Nearest Neighbors technique. *Remote Sensing of Environment*, *82*, 457–468.

McRoberts, R. E., Tomppo, E. O., & Næsset, E. (2010). Advances and emerging issues in national forest inventories. *Scandinavian Journal of Forest Research*, *26*, 368–381.

McRoberts, R. E., Tomppo, E. O., Finley, A. O., & Heikkinen, J. (2007). Estimating areal means and variances using the k-Nearest Neighbors technique and satellite imagery. *Remote Sensing of Environment*, *111*, 466–480.

McRoberts, R. E., Bechtold, W. A. B., 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*, 304–308.

Moeur, M., & Stage, A. R. (1995). Most similar neighbor — An improved sampling inference procedure for natural resource planning. *Forest Science*, *41*(2), 337–359.

Nothdurft, A., Saborowski, J., & Breidenbach, J. (2009). Spatial prediction of forest stand variables. *European Journal of Forest Research*, *128*, 241–251.

Ohmann, J. L., & Gregory, M. J. (2002). Predictive mapping of forest composition and structure with direct gradient analysis and nearest neighbor imputation in coastal Oregon, U.S.A. *Canadian Journal of Forest Research*, *32*, 725–741.

Quenouille, M. (1949). Approximate tests of correlations in time series. *Journal of the Royal Statistical Society, Series B*, *11*, 18–44.

Rao, J. N. K. (2007). Jackknife and bootstrap methods for small area estimation. *Proceedings of the survey research methods section, American Statistical Association* Available at. http://www.amstat.org/Sections/Srms/Proceedings/y2007/Files/JSM2007-000358.pdf (last accessed: April 2011).

Royall, R. M., & Herson, J. (1973). Robust estimation in finite populations II. *Journal of the American Statistical Association*, *68*(344), 890–893.

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.

Särndal, C.-E., Swensson, B., & Wretman, J. (1992). *Model assisted survey sampling.* New York: Springer 694 p.

Shao, J. (1996). Resampling methods in sample surveys. *Statistics*, *27*, 203–254.

Shao, J., & Sitter, R. R. (1996). Bootstrap for imputed survey data. *Journal of the American Statistical Association*, *91*(435), 1278–1288.

Simpson, J. A., & Weiner, E. S. C. (1989). (Second Edition). *The Oxford English Dictionary*, *Vol. 7*. (pp. 923–924)Oxford: Clarendon Press Preparers.

Steele, B. M., & Patterson, D. A. (2000). Ideal bootstrap estimation of expected prediction error for k-nearest neighbor classifiers: Applications for classification and error assessment. *Statistics and Computing*, *10*, 349–355.

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

Tomppo, E. O., 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.

Tukey, J. (1958). Bias and confidence in not quite large samples. *Annals of Mathematical Statistics*, *29*, 614.