Forest Ecology and Management 272 (2012) 3–12

Contents lists available at SciVerse ScienceDirect

Forest Ecology and Management

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

Estimating forest attribute parameters for small areas using nearest neighbors techniques

Ronald E. McRoberts *

Northern Research Station, US Forest Service, 1992 Folwell Avenue, Saint Paul, MN 55108, USA

ARTICLE INFO

Article history:
Available online 3 September 2011

Keywords:
Optimization
Distance metric
Neighbor weighting
k-value
Variance
Diagnostics

ABSTRACT

Nearest neighbors techniques have become extremely popular, particularly for use with forest inventory data. With these techniques, a population unit prediction is calculated as a linear combination of observations for a selected number of population units in a sample that are most similar, or nearest, in a space of ancillary variables to the population unit requiring the prediction. Nearest neighbors techniques are appealing for multiple reasons: they can be used with categorical response variables for which the objective is classification and with continuous response variables for which the objective is prediction; they can be used for both univariate and multivariate prediction; they are non-parametric in the sense that no assumptions regarding the distributions of response or predictor variables are necessary; they are synthetic in the sense that they can readily use information external to the geographic area for which an estimate is sought; they are useful for map construction, small area estimation, and inference; and they can be used with a wide variety of data sets. Recent advances and emerging issues in nearest neighbors techniques are reviewed for four topic areas: (1) distance metrics, (2) optimization, (3) diagnostic tools, and (4) inference. The focus of the study is estimation of mean forest stem volume per unit area for small areas using a combination of forest inventory observations and Landsat Thematic Mapper (TM) imagery. However, the concepts and techniques are generally applicable for all nearest neighbors problems.

Published by Elsevier B.V.

1. Introduction

Among techniques that have been investigated for predicting forest attributes from satellite imagery and ground data, nearest neighbors techniques have enjoyed particular popularity within the forest inventory community. With these techniques, 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 ancillary variables to the unit requiring a prediction. Nearest neighbors techniques are appealing for multiple reasons: (1) they can be used with categorical response variables for which the objective is classification and with continuous response variables for which the objective is prediction, (2) they can be used for both univariate and multivariate prediction, (3) they are non-parametric in the sense that no assumptions regarding the distributions of response or predictor variables are necessary, (4) they are synthetic in the sense that they can readily use information external to the geographic area for which an estimate is sought, (5) they are useful for map construction, small area estimation, and inference, and (6) they can be used with a wide variety of data sets.

Nearest neighbors techniques were first introduced in an unpublished US Air Force report by Fix and Hodges (1951) as a non-parametric discriminant technique for classification into populations whose distributions are unknown. This seminal paper was republished in Agarwal (1977) and as Fix and Hodges (1989). In a commentary on Fix and Hodges (1951), Silverman and Jones (1989) assert that the popularity of nearest neighbors techniques can be attributed to Cover and Hart (1967) who showed that the classification error rate when the technique is used with categorical response variables has an upper bound. Further, they assert that das Gupta (1973) was the first to acknowledge fully "the breadth of ideas" expressed in Fix and Hodges (1951).

Much of the early foundational work on the use of nearest neighbors techniques for classification purposes appears in the pattern recognition and machine learning literature. Within the natural resources area, forestry applications have increased dramatically following seminal papers by Tomppo (1991, 1996) and Moeur and Stage (1995). Forestry applications have truly been international with published reports for Austria (Koukal et al., 2007), Canada (LeMay and Temesgen, 2005), China (Tomppo et al., 2001; Xu et al., 2011), Costa Rico (Thessler et al., 2008), Ecuador (Rajaniemi et al., 2005), Finland (Tomppo, 2006; Tomppo and Halme, 2004), Germany (Nothdurft et al., 2009; Breidenback et al., 2010), Ireland (McInerney and Nieuwenhuis, 2009), Italy

* Tel.: +1 651 649 5174; fax: +1 651 649 5140.
E-mail address: rmcroberts@fs.fed.us

0378-1127/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.foreco.2011.06.039

(Maselli et al., 2005; Baffetta et al., 2009, Chirici et al., 2008), Japan (Kajisa et al., 2008), Korea (Kim et al., 2011), New Zealand (Tomppo et al., 1999), Norway (Gjertsen, 2007), Peru (Salovaara et al., 2005), Siberia (Fuchs et al., 2009), Sweden (Holmström and Fransson, 2003, Nilsson et al., 2005, Reese et al., 2003, the United States of America (USA) (Moeur and Stage, 1995; Ohmann and Gregory, 2002; McRoberts et al., 2002, 2007), and Zambia (Maltamo and Eerikäinen, 2001). A bibliography of published peer-reviewed, nearest neighbors papers is available at: http://blue.for.msu.edu/NAFIS/biblio.html (last accessed: June 2011).

The primary forest inventory applications of nearest neighbors techniques are in four areas: (1) imputation of missing values for forest inventory and monitoring databases (LeMay and Temesgen, 2005; LeMay et al., 2008; Moeur and Stage, 1995; Temesgen et al., 2003, 2008; Eskelson et al., 2009), (2) mapping (Franco-Lopez et al., 2001; McRoberts et al., 2002, 2007; Koukal et al., 2007; Ohmann et al., 2011; Chirici et al., 2008; Tomppo et al., 2008), (3) small area estimation (Tomppo, 1996; McRoberts et al., 2007), and (4) support for probability-based and model-based inference (Baffetta et al., 2009, 2011; Katila and Tomppo, 2005; McRoberts et al., 2002, 2007; Nilsson et al., 2005; Tomppo, 1991, 1996; Nilsson et al., 2005; Magnussen et al., 2009, 2010a,b).

Recent forestry-related investigations have shifted from simple descriptions of nearest neighbors applications to more foundational work on efficiency and inference. McRoberts (2009) reported diagnostic tools for evaluating and enhancing nearest neighbors prediction for continuous, univariate response variables. Finley et al. (2006) and Finley and McRoberts (2008) investigated enhanced search algorithms for identifying nearest neighbors. Tomppo and Halme (2004), McRoberts (2008), and Tomppo et al. (2009) used a genetic algorithm approach to optimize a matrix-based distance metric. Magnussen et al. (2010b) developed an approach for dealing with extrapolation problems. Baffetta et al. (2009, 2011) demonstrated use of nearest neighbors techniques with a model-assisted approach to inference. Multiple approaches to variance estimation have been reported: Kim and Tomppo (2006) reported a method for estimating the uncertainty of predictions using a variogram approach; McRoberts et al. (2007) derived a model-based k-NN variance estimator; Magnussen et al. (2009) investigated an estimator of mean square error; and Nothdurft et al. (2009) and Magnussen et al. (2010a) used resampling approaches.

The objective of the study was twofold. First, a review of the current state of nearest neighbors techniques is provided for four topic areas: (1) distance metrics, (2) optimization with respect to neighbor weighting and selection of the number of nearest neighbors, (3) diagnostics, and (4) inferential methods. Second, for each topic area, an analysis was conducted to highlight a recent finding and/or to suggest additional directions for relevant research. The analyses were constrained to small area inference for the continuous response variable, forest stem volume, per unit area for multiple reasons: first, many of the issues associated with categorical response variables have been addressed in the pattern recognition and machine learning literature; second, Eskelson et al. (2009) provide a good review of the use of nearest neighbors techniques for imputation purposes; and third, small area estimation and inference subsume issues related to mapping. Finally, the examples all use satellite imagery as ancillary information because of the popularity of such applications.

2. Data

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) (Kauth and Thomas, 1976; Crist and Cicone, 1984) for each of the three image dates were used. Within the study area, four 4-km $\times$ 4-km areas of interest (AOI) were also selected. These AOIs represent small areas in an inventory context in the sense that the number of plots within the AOIs is insufficient to produce acceptably precise estimates for most forest attributes.

Data were obtained for plots established by the Forest Inventory and Analysis (FIA) program of the US Forest Service which conducts the national forest inventory of the USA (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. In general, centers of forested, partially forested, or previously forested plots are determined using global positioning system (GPS) receivers, whereas centers of non-forested plots are verified using aerial imagery and digitization methods. Field crews observe species and measure diameter at breast height (dbh, 1.37 m, 4.5 ft) and height for all trees with dbh $\geqslant$ 12.7 cm (5 inch). These data and statistical models are used to estimate individual tree volumes which are aggregated to obtain subplot-level volume estimates (m^3/ha). Subplot-level volume estimates were combined with the spectral data for pixels containing subplot centers. Data were obtained between 1999 and 2003 for 2266 plots in the study area, but data for only the central subplot of each plot were used for this study to avoid issues of spatial correlation among subplot observations.

3. Nearest neighbors techniques

3.1. Notation and terminology

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 ancillary variables with observations for all population units. In the terminology of nearest neighbors techniques, the ancillary 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 complete sets of observations for all feature variables.

For continuous response variables, the nearest neighbors prediction, $\tilde{y}_i$, for the ith target set element is calculated as,

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

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 (Section 3.2), and w_{ij} is the weight assigned to the jth nearest neighbor with $\sum_{j=1}^{k} w_{ij} = 1$ (Section 3.3.1). For categorical variables the predicted class of the ith target set element is the most heavily weighted class among the k nearest neighbors, a weighted median or mode in case of ordinal scale variables, and a mode in the case of nominal variables.

The term k nearest neighbors (k-NN) is generic and may be used to refer to any nearest neighbors technique. Implementation of k-NN requires three primary selections: (1) a distance metric, (2) a

neighbor weighting scheme, and (3) a value for k. These selections are often guided by assessments of results obtained for various combinations, and the assessments, in turn, rely on diagnostics related to the quality of predictions, analysis of residuals, extrapolations, and heteroscedasticity.

3.2. Distance metrics

Distance metrics may be categorized with respect to multiple factors: (1) local or global, (2) independent of or dependent on response variable observations, and (3) whether they can be expressed in matrix form. Although only global metrics that are uniform throughout feature space are considered for this study, local distance metrics that vary depending on the region of feature space have also been proposed (Friedman, 1994; Ricci and Avesani, 1996; Hastie and Tibshirani, 1996). Metrics that are independent of response variable observations include the *Minkowsky* family of metrics (Kruskal, 1964),

$$d_{ij} = \left(\sum_{m=1}^{M} |x_{im} - x_{jm}|^p \right)^{\frac{1}{p}},$$

where m indexes feature variables and p is a parameter whose value is to be selected; *Manhattan* or *City Block* ($p = 1$) and *Euclidean* ($p = 2$) metrics are special cases of the Minkowsky metric. Metrics that depend on observations of the response variable include the *Canonical Correlation Analysis* metric (Moeur and Stage, 1995; LeMay and Temesgen, 2005; LeMay et al., 2008; Temesgen et al., 2003, 2008; Maltamo and Eerikäinen, 2001), the *Canonical Correspondence Analysis* metric (Ohmann and Gregory, 2002; Pierce et al., 2009; Ohmann et al., 2011), and the *Fuzzy*, *Multiple Regression*, and *Non-parametric* metrics (Maselli et al., 2005; Chirici et al., 2008).

Many familiar metrics can be expressed in matrix form as,

$$d_{ij}^2 = (\mathbf{X}_i - \mathbf{X}_j)' \mathbf{M} (\mathbf{X}_i - \mathbf{X}_j), \quad (2)$$

where i denotes a target set element for which a prediction is sought, j denotes a reference set element, $\mathbf{X}_i$ and $\mathbf{X}_j$ are vectors of observations of feature variables for the ith and jth elements, respectively, and $\mathbf{M}$ is a square, positive definite 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. Metrics based on canonical correlation and canonical correspondence analyses can also be expressed in matrix form.

Despite the progress in many areas related to nearest neighbors techniques, few comparisons of distance metrics or assessments of the utility of attempting to optimize distance metrics have been reported. For predicting forest stand attributes using variables obtained from aerial photography, LeMay and Temesgen (2005) reported that a metric based on canonical correlation analysis was superior to Euclidean distance and Manhattan distance. For predicting forest attributes from Landsat-based variables, Chirici et al. (2008) reported that a distance metric giving greater weights to reference pixels whose response variable observations are closer to the mean of the observations was superior to Euclidean and Mahalanobis metrics. Tomppo and Halme (2004), McRoberts (2008), and Tomppo et al. (2009) used a genetic algorithm (Holland, 1975) approach to optimize a weighted Euclidean distance metric. The conclusion is that although a variety of metrics have been reported, the number of reported comparisons is small and no consistent conclusions have been forthcoming. In addition, the benefits of optimizing a distance metric have not been rigorously investigated, either in terms of optimizing a criterion in the reference set or evaluating the degree to which optimization in the reference set is realized in the target set. Thus, a more comprehensive investigation of nearest neighbor distance metrics is warranted.

3.3. Optimization

3.3.1. Neighbor weighting

The most common approach to weighting neighbors in the calculation of predictions is to weight the jth nearest neighbor reference element inversely proportional to its distance, d_{ij}, to the ith target element, i.e.,

$$w_{ij} \propto d_{ij}^{-t},$$

where $t \geqslant 0$. Although commonly selected values are $t = 0$, $t = 1$, and $t = 2$, no investigations of comparisons of the effects of these values of t or attempts to optimize the selection are known to have been reported.

3.3.2. Value of k

The value of k may be selected to optimize multiple criteria either individually or in combination. For some approaches to nearest neighbors prediction such as Most Similar Neighbor (MSN) (Moeur and Stage, 1995) which uses the canonical correlation analysis metric and Gradient Nearest Neighbor (GNN) (Ohmann and Gregory, 2002; Pierce et al., 2007; Ohmann et al., 2011) which uses the canonical correspondence analysis metric, $k = 1$ is always selected. For approaches that permit $k > 1$, smaller values of k are generally preferred as a means of reducing computational intensity. However, caution must be exercised when selecting small values of k because such values may yield root mean square errors (RMSE) that are greater than the standard deviation of the response variable observations, meaning that the overall mean as a prediction for every target element is better at minimizing RMSE than are the k-NN predictions.

The most intuitive approach to selecting k is to determine the value that optimizes a criterion such as RMSE for the reference set. When RMSE is large and for large reference sets, the value of k that minimizes RMSE is also often large. However, the RMSE versus k curve is often relatively flat in the vicinity of the optimal value of k so that a smaller value of k may be selected with little impact on RMSE. In fact, McRoberts et al. (2002) proposed selecting the smallest value of k for which RMSE is not more than a predetermined percentage (e.g., 1% or 5%) greater than the smallest value of RMSE.

Other optimization criteria may also be considered. For example, k-NN predictions corresponding to extreme observations of the response variable are often consistently smaller or consistently greater than those extreme observations. Thus, the value of k that minimizes this under- or over-estimation problem could be selected. For multivariate problems, a criterion such as $T = \sum_{p=1}^{P} w_p T_p^2$ could be used where p indexes response variables, $\{w_p\}$ is a set of arbitrarily selected weights that sum to 1, $T_p^2 = \frac{SS_{mean} - SS_{res}}{SS_{mean}}$, $SS_{mean} = \sum_{i=1}^{n} (y_{pi} - \bar{y}_p)^2$, and $SS_{res} = \sum_{i=1}^{n} (y_{pi} - \hat{y}_{pi})^2$. T_p^2 is similar to the familiar R^2 that is used to assess quality of fit for regression models; a different notation is used because T_p^2 does not satisfy all the criteria for R^2 (Anderson-Sprecher, 1994). Selection of the value of k is often conducted in the reference set under the assumption that optimization in the reference set will produce at least near optimal results in the target set. However, optimization could be conducted directly using the target set. For example, the value of k could be selected to minimize the variance of the estimate of the mean of the population represented by the target set (Section 3.5). The important point is that multiple optimization criteria may be used to select values of k. Because the various

criteria do not necessarily lead to similar values of k, selection of the value of k may represent a compromise which yields sub-optimal results for all criteria. However, the magnitudes of the necessary compromises are generally unknown.

3.4. Diagnostics

McRoberts (2009) reported k-NN diagnostic tools for univariate response variables. Of these tools, techniques for assessing lack of fit and for identifying outliers and influential observations are particularly relevant. An important concern with most approaches to nearest neighbors is that predictions corresponding to the smallest observations are overestimated because, of necessity, observations for all nearest neighbors are larger than that smallest observation. Similarly, predictions corresponding to the largest observations are underestimated. The degree to which this lack of fit phenomenon affects overall estimation should not be ignored.

Outliers are defined as observations that differ from other observations to such a degree that they raise questions as to whether they are from a different population or whether the sampling is faulty (Kendall and Buckland, 1982). The standard approach for identifying outliers is to calculate standardized residuals as the ratios of residuals and their standard deviations and then to assess the probability of observing the standardized residuals under specified distributional assumptions. Assuming standardized residuals follow a Gaussian (0,1) distribution, the proportion of standardized residuals with absolute values greater than 2.0 and 3.0 should be less than approximately 0.045 and 0.001, respectively. However, when combining data from different sources, multiple factors contribute to the potential for greater proportions of observations being characterized as outliers than would be expected for data from a single source. First, plots with similar total tree volumes but different age structures or health conditions may have quite different spectral signatures. Second, the observation for a smaller plot may not adequately characterize a larger pixel. Third, disturbance on the plot between the plot observation and image acquisition dates is not uncommon. Fourth, correct registration of the plot coordinate system to the satellite image coordinate system is highly dependent on the quality of geographic positioning system receivers (McRoberts, 2010b).

Observations that cause substantial changes in predictions are further characterized as influential observations. For nearest neighbors techniques, the influence of individual reference set observations may be partially assessed using the sums of weights used to calculate predictions for either all reference set units or all target set units. Potential influential reference set observations may be identified by graphing standardized residuals against sums of weights for the corresponding reference set observations. The effects of reference observations with large absolute values of standardized residuals whose sums of weights are also large should be assessed by comparing RMSEs or other statistics obtained by including and excluding such reference observations. Formal statistical tests for detecting influential observations for parametric models may be found in Chatterjee and Yilmaz (1992) and Belsey et al. (1980), but few references are available for non-parametric approaches.

3.5. Inference

In a sampling framework, inference requires expression of the relationship between a population parameter, μ, and its estimate, $\hat{\mu}$, in probabilistic terms (Dawid, 1983). These probabilistic expressions often take the form of confidence intervals,

$$\hat{\mu} \pm t_{1-\alpha}\sqrt{\hat{\mathrm{Var}}(\hat{\mu})} \tag{3}$$

where $1-\alpha$ denotes the probability that confidence intervals constructed using data for all possible samples will include μ. Of crucial importance, construction of the confidence intervals, and therefore inference, requires estimates, $\hat{\mu}$ and $\hat{\mathrm{Var}}(\hat{\mu})$.

Two approaches to inference are common, probability-based (design-based) and model-based. McRoberts (2010a) summarizes assumptions underlying probability-based inference. Baffetta et al. (2009, 2011) illustrate probability-based inference based on the model-assisted difference estimator with the k-NN technique for use with categorical response variables. However, as with all probability-based estimators, model-assisted estimators often suffer the effects of small sample sizes when used for small area estimation.

The remainder of this study focuses on model-based inference because of its utility for small area estimation. With model-based inference, as with probability-based model-assisted inference, a modeling procedure is used to produce predictions for population units. Model-based approaches to inference rest on two primary assumptions: (1) each observation is a realization from an entire distribution of possible observations for each population unit, not just a constant value as is the case for probability-based inference; and (2) randomization enters through the realization of observations from the distributions for individual population units selected for the sample, not from the random selection of population units into a sample. Thus, model-based approaches do not require probability samples. Although model-based estimators are often more computationally intense, they produce maps as by-products; they produce estimates that are compatible with maps; they may produce viable estimates for small areas; but they cannot be assumed to be unbiased. A consequence of the latter feature is that separate assessments of model quality of fit are often required for model-based inference.

With model-based inference, the mean and standard deviation of the distribution of $\mathbf{Y}$ for the ith population unit may be denoted μ_i and σ_i, respectively, and an observation for the ith population unit may be 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. With model-based approaches, estimation often focuses on μ_i rather than the particular observation which is a random realization from the distribution of which μ_i is the mean. With nearest neighbors techniques, the estimator of μ_i is $\hat{\mu}_i = \tilde{y}_i$ from Eq. (1), and the estimator of the population mean is,

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

McRoberts et al. (2007) derived parametric estimators for σ_i^2 and $\mathrm{Var}(\hat{\mu})$ that accommodate spatial correlation among reference set observations. For the inference portion of the study, observations of only the central subplots of FIA plots were used, so that spatial correlation could be ignored. In the absence of spatial correlation, σ_i^2 may be estimated as,

$$\hat{\sigma}_i^2 = \frac{\sum_{j=1}^{k}\left(y_j^i - \hat{\mu}_i\right)^2}{k-1}, \tag{6}$$

where $\{y_j^i\}$ are as defined for Eq. (1). The general form of a model-based estimator of the variance of $\hat{\mu}$ is,

$$\mathrm{Var}(\hat{\mu}) = \frac{1}{N^2}\left[\sum_{i=1}^{N}\mathrm{Var}(\hat{\mu}_i) + 2\sum_{i<j}^{N}\sum^{N}\mathrm{Cov}(\hat{\mu}_i, \hat{\mu}_j)\right]. \tag{7}$$

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

$$\hat{\text{Cov}}(\hat{\mu}_i, \hat{\mu}_j) \approx \frac{m_{ij}\hat{\sigma}_i\hat{\sigma}_j}{k^2}, \quad (8)$$

where m_{ij} is number of nearest neighbors common to both the ith and jth population units. $\text{Cov}(\hat{\mu}_i, \hat{\mu}_j)$ may be estimated by substituting $\hat{\sigma}_i$ from Eq. (6) into Eq. (8). Further, if the ith and jth population units share no common nearest neighbors, $\text{Cov}(\hat{\mu}_i, \hat{\mu}_j) = 0$. The variance estimate for the estimate of the ith mean, $\hat{\mu}_i$, is simply the special case of Eq. (8) for which $i = j$, so that,

$$\hat{\text{Var}}(\hat{\mu}_i) \approx \frac{\hat{\sigma}_i^2}{k}. \quad (9)$$

Thus, $\text{Var}(\hat{\mu})$ may be estimated by substituting from Eqs. (8) and (9) into Eq. (7).

As is apparent, this parametric approach to variance estimation is complex, and because Eq. (7) requires double summations over all population units, the approach is also computationally intensive. In addition, the effects of small values of k have not been investigated, particularly for Eq. (6). Thus, an alternative variance estimator such as the bootstrap merits consideration.

4. Methods

For future reference, the terms satellite image *pixel*, population *unit*, and reference and target set *element* are used interchangeably.

4.1. Dimension reduction

Some aspects of nearest neighbors techniques, such as concern for lack of fit, outliers, and influential reference observations, are common with other estimation techniques, but other aspects are unique to nearest neighbors techniques. One of the latter aspects is that inclusion of feature variables that are unrelated to response variables produces detrimental effects for many distance metrics, whereas for linear regression the effects of such unrelated variables on predictions are mitigated by corresponding coefficient estimates that are near zero. Therefore, for nearest neighbors techniques reduction of the dimension of feature space may have positive consequences. For this study, all combinations of all numbers of feature variables were compared with respect to RMSE using the Euclidean distance metric, equal weighting of neighbors, and the leave-one-out method. The number and combination corresponding to the smallest RMSE were selected for most subsequent analyses. Unless otherwise indicated, the best combination of six variables was used because it produced the overall minimum RMSE (Table 1).

4.2. Distance metrics

Euclidean distance is the most intuitive metric, and most other metrics represent attempts to improve predictions. However, the degree to which other metrics actually improve predictions is generally unknown, and the degree to which optimization in the reference set is realized in the target set is also generally unknown. Therefore, a study was conducted to assess the degree to which optimization of a matrix-based distance metric in the reference set improves predictions in an independent target set over predictions obtained using the Euclidean metric. Because of computational intensity, these analyses used only the best combination of three feature variables (Table 1). Further, the analyses used equal neighbor weighting ($t = 0$).

The reference data for the study area were fit using a polynomial model that included an intercept and coefficients corresponding to all three feature variables, their squares, and the products of all pairs of the three variables. Simulated data sets were constructed using three steps. First, volume, $\hat{y}$, for each reference set element was predicted using the polynomial model and the observed values of the three feature variables. Second, residuals, ε, were randomly selected from a Gaussian (0,1) distribution, but truncated to $|\varepsilon| \leqslant 2.5$. Third, observations for eight data sets were simulated as $\hat{y} + c\varepsilon$ where $c = 0.025$, 0.05, 0.10, 0.15, 0.25, 0.50, 1.00, and 2.00.

Each simulated data set was randomly divided into a reference set and a target set of equal size. Three matrix-based distance metrics as per Eq. (2) were used: (1) a Euclidean distance metric with matrix, **M**, expressed as,

$$\mathbf{M}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \quad (10)$$

(2) a weighted Euclidean distance metric with matrix, **M**, expressed as,

$$\mathbf{M}_2 = \begin{pmatrix} m_{11} & 0 & 0 \\ 0 & m_{22} & 0 \\ 0 & 0 & m_{33} \end{pmatrix}, \quad (11)$$

where $m_{ii} > 0$; and (3) a distance metric with matrix, **M**, expressed as,

$$\mathbf{M}_3 = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix}, \quad (12)$$

where $m_{12} = m_{21}$, $m_{13} = m_{31}$, $m_{23} = m_{32}$, and $|\mathbf{M}_3| > 0$.

Table 1
Best combinations of feature variables.[a]

Number of variables	Best combination	Optimal		$\text{Optimal}_{1.01}$		$\text{Optimal}_{1.05}$	
		k	RMSE	k	RMSE	k	RMSE
1	6	169	67.99	62	68.33	16	69.57
2	4-11	111	65.43	37	65.68	16	67.04
3	1-6-11	76	65.08	56	65.39	17	66.68
4	1-6-10-11	47	64.96	25	65.15	11	66.50
5	1-6-7-10-11	26	64.85	16	65.06	10	66.28
6	1-6-7-8-10-11	25	64.73	18	65.03	12	65.86
7	1-2-4-6-8-10-11	30	64.96	22	65.27	11	66.48
8	2-3-4-6-8-9-10-11	52	65.08	37	65.36	11	66.63
9	1-2-3-4-6-8-9-10-11	38	65.13	33	65.39	14	66.67
10	1-2-3-4-6-7-8-9-10-11	44	65.29	24	65.60	10	66.90
11	1-2-3-4-6-7-8-9-10-11-12	60	65.49	29	65.77	12	66.94
12	1-2-3-4-5-6-7-8-9-10-11-12	63	65.58	31	65.87	13	66.99

[a] Optimal: Minimizes sum of squared residuals (SSres). $\text{Optimal}_{1.01}$: Smallest k value such that $\text{SSres} \geqslant 1.01 \cdot \text{SSres}_{\text{opt}}$. $\text{Optimal}_{1.05}$: Smallest k value such that $\text{SSres} \geqslant 1.05 \cdot \text{SSres}_{\text{opt}}$.

For each of the eight values of c used to generate residuals, the value of k that minimized RMSE for the reference set using the metric based on the $\mathbf{M}_1$ matrix and a leave-one-out approach was determined. In addition, for each value of c, the approximate values of the elements of $\mathbf{M}_2$ and $\mathbf{M}_3$ and the value of k that minimized RMSE for the reference set using a leave-one-out approach were determined. A grid search technique was used whereby the space of possible values for matrix elements was searched using a systematic grid with small distances between grid lines. The RMSEs were denoted $RMSE_{ref,M_j}$ where the subscript M_j denotes the distance metric based on the matrix $\mathbf{M}_j$. The degrees to which metrics based on $\mathbf{M}_2$ and $\mathbf{M}_3$ were superior to the metric based on $\mathbf{M}_1$ in the reference set were estimated using the ratios $\frac{\text{RMSE}_{\text{ref},M_2}}{\text{RMSE}_{\text{ref},M_1}}$ and $\frac{\text{RMSE}_{\text{ref},M_3}}{\text{RMSE}_{\text{ref},M_1}}$.

For each value of c, distance metrics based on the three matrices and their corresponding optimal values of k obtained for the reference set were applied to the target set. RMSEs for the target set were calculated using sums of squared differences between target set observations and predictions calculated using nearest neighbors selected from the reference set. The RMSEs were denoted $RMSE_{\text{tgt},M_j}$ where the subscript M_j denotes the distance metric based on $\mathbf{M}_j$. The degree to which metrics based on optimized $\mathbf{M}_2$ and $\mathbf{M}_3$ matrices in the reference set were superior to the Euclidean metric based on the $\mathbf{M}_1$ matrix in the target set was estimated using the ratios $\frac{\text{RMSE}_{\text{tgt},M_2}}{\text{RMSE}_{\text{tgt},M_1}}$ and $\frac{\text{RMSE}_{\text{tgt},M_3}}{\text{RMSE}_{\text{tgt},M_1}}$. Analyses for each of the eight simulated data sets, one for each value of c, were replicated 25 times.

For each replication for each data set, $T^2 = \frac{\text{SS}_{\text{mean}} - \text{SS}_{\text{res},M_1}}{\text{SS}_{\text{mean}}}$ was calculated for the simulated reference set where SS_{mean} and SS_{res,M_1} are as defined in Section 3.3.2. The mean value of T^2 over the 25 replications was calculated for each value of c, and the means of the ratios of RMSEs were graphed against corresponding means of T^2. In addition, the T^2 value was calculated for the observed reference set.

4.3. Optimizing neighbor weighting and the value of k

The most common approach to neighbor weighting is to calculate weights for Eq. (1) as,

$$w_{ij} = \frac{d_{ij}^{-t}}{W_i} \tag{13}$$

where d_{ij} is the distance from the jth reference set element to the ith target set element, $t \geqslant 0$, and $W_i = \sum_{j=1}^{k} d_{ij}^{-t}$. Values of k and t that minimized RMSE for the Euclidean distance metric using the approach described by Eq. (13) were determined. In addition, values of k that minimized RMSE for the Euclidean distance metric were determined using the same weighting except with the commonly used values of t, i.e., $t = 0$, $t = 1$, and $t = 2$. A graph of RMSEs corresponding to $t = 0$, $t = 1$, $t = 2$, and the optimal values for all values of k were used to assess the utility of determining an optimal value of t.

As noted in Section 3.3.2, the value of k may be selected to optimize any of multiple criteria of which RMSE calculated for the reference set is the most familiar and commonly used. However, an optimization criterion for the target set could be minimization of $\hat{\text{Var}}(\hat{\mu})$ where $\hat{\mu}$ is the estimated mean of the AOI and $\hat{\text{Var}}(\hat{\mu})$ is obtained using Eq. (7). A graph of RMSE for the reference set and $\sqrt{\hat{\text{Var}}(\hat{\mu})}$ for the four AOIs versus k illustrates differences in the optimal value of k corresponding to different optimization criteria.

4.4. Diagnostics

For the best combination of six feature variables, the associated value of k (Table 1), the Euclidean distance metric, and equal neighbor weighting ($t = 0$), k-NN predictions were calculated. In addition, residuals were calculated as differences between response variable observations and predictions. The observations, predictions, and residuals were jointly and simultaneously ordered by values of the predictions and aggregated into groups of size 50. For each group, the means of the observations and predictions and the standard deviation of the residuals were calculated. Lack of fit was assessed by graphing group observation means against group prediction means. In addition, a nonlinear model of the form,

$$y = \beta_1[1 - \exp(\beta_2 X)] + \varepsilon,$$

was fit to the group standard deviations using the group means of predictions as the independent variable. Standardized residuals for individual reference set elements were calculated as ratios of observed residuals and the standard deviation estimates corresponding to the elements' k-NN predictions and were graphed against corresponding sums of weights. Potential influential observations, identified as reference observations with large standardized residuals and large sums of weights, were then assessed by comparing RMSE for the entire reference set to RMSEs for reference sets that excluded the potential influential observations.

4.5. Inference

The parametric approach to variance estimation is complex and computationally intensive, and in addition the underlying assumptions have not been rigorously investigated. Resampling procedures such as the bootstrap are well-suited for estimation for complex and non-parametric model applications and for applications requiring assumptions whose validity is difficult to assess. The bootstrap resampling procedure was invented by Efron (1979, 1981, 1982) and further developed by Efron and Tibshirani (1994). All bootstrap methods depend on the notion of a *bootstrap sample*.

For modeling problems, Efron and Tibshirani (1994) describe an approach to bootstrapping characterized as *bootstrapping pairs*. With this approach, the bootstrap sample consists of a sample of n pairs $(y_i, \mathbf{X}_i)$ that is drawn with replacement from the reference set. For each bootstrap sample, b, the nearest neighbors technique is used to calculate predictions for each population unit, and the population estimate, $\hat{\mu}_{boot}^b$, is then calculated using Eq. (5). The overall bootstrap population estimate is then,

$$\hat{\mu}_{\text{boot}} = \frac{1}{B}\sum_{b=1}^{B} \hat{\mu}_{\text{boot}}^b, \tag{14}$$

where B is the number of bootstrap samples. The bootstrap estimate of bias is defined as,

$$\hat{\text{Bias}}_{\text{boot}}(\hat{\mu}) = \hat{\mu}_{\text{boot}} - \hat{\mu} \tag{15}$$

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

$$\hat{\text{Var}}_{\text{boot}}(\hat{\mu}) = \frac{1}{B-1}\sum_{b=1}^{B} \left(\hat{\mu}_{\text{boot}}^b - \hat{\mu}_{\text{boot}}\right)^2. \tag{16}$$

Bootstrap resampling continued until estimates of both $\hat{\mu}_{\text{boot}}$ and $\hat{\text{Var}}_{\text{boot}}(\hat{\mu})$ stabilized. Variance estimates obtained using the bootstrapping pairs approach were compared to estimates obtained using the parametric approach as described in Eqs. (5)–(9).

5. Results and discussion

5.1. Distance metrics

Comparisons of the Euclidean distance metric and metrics based on optimized diagonal and full matrices with respect to RMSE in the target set indicate that unless T^2 (Section 4.2) for the Euclidean metric in the reference set is large, optimization produces very little benefit (Fig. 1). In fact, for small T^2 the Euclidean distance metric was superior to the metrics based on the optimized matrices. The latter result is attributed to optimization with respect to random variation in the reference set observations rather than to underlying relationships between the response and feature variables. This phenomenon is similar to overfitting for a regression model. For the actual reference set observations, $T^2 \approx 0.20$ which is in the range for which the Euclidean metric would be expected to produce estimates as good as an optimized matrix-based distance metric.

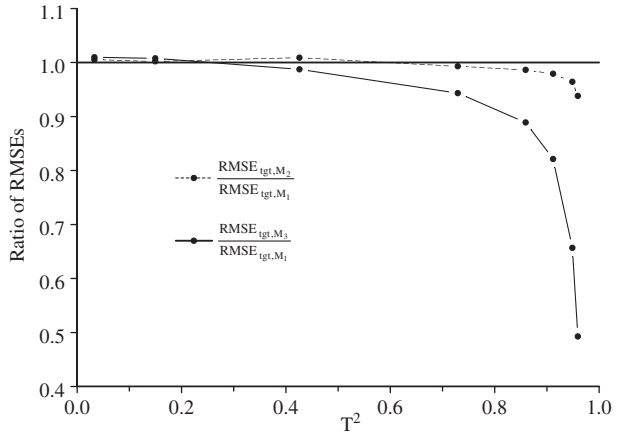

Fig. 1. Effects of optimizing matrix-based distance metrics.

5.2. Optimizing neighbor weighting and the value of k

The beneficial effects of selection of an optimal value of t for neighbor weighting were minimal. For reference set analyses for the best combination of six feature variables and the Euclidean distance metric, $t_{opt} = 0.71$ and $k = 25$ produced the smallest RMSE. However, the RMSEs for these optimal combinations of t and k were not substantially smaller than the RMSEs for $t = 0$, $t = 1$, and $t = 2$ with their corresponding optimal values of k (Fig. 2). Larger values of t produced smaller RMSEs regardless of the value of k, although the differences in RMSEs were not substantial. Finally, increasing values of t_{opt} corresponding to values of k increasing beyond the optimal combination of t and k produced a compensating phenomenon. For a given target set element, the distance to a reference set neighbor increases as the order or ranking of the neighbor increases, i.e., the first neighbor is closer to the target set element than are subsequent neighbors. Therefore, as k increases, the distance from the kth neighbor to the target set element also increases. However, because t_{opt} also increases as k increases (Fig. 3), the relative weight for the kth neighbor decreases. The combined result is that as k increases from smaller to larger values, the individual contributions of additional neighbors to the k-NN prediction decreases to 0. This phenomenon is reflected in the relative flatness of the RMSE versus k curve for t_{opt} as k increases beyond its optimal value (Fig. 2).

Regardless of the value of t, small values of k produced RMSEs that were greater than the sum of squared deviations of observations around their mean Fig. 2. Thus, caution is advised when selecting small values of k as a means of reducing computational intensity.

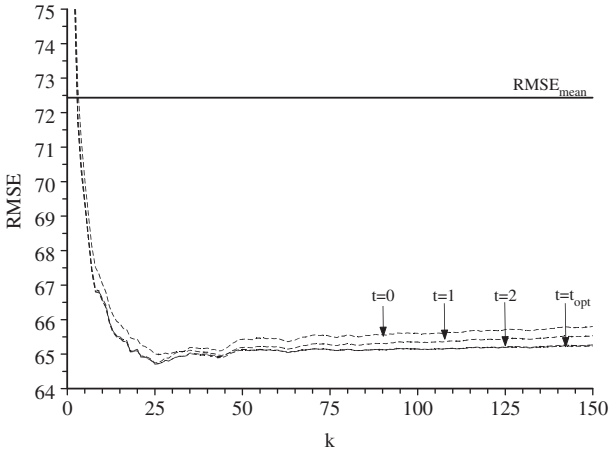

Fig. 2. The effects of k and t on root mean square error (RMSE).

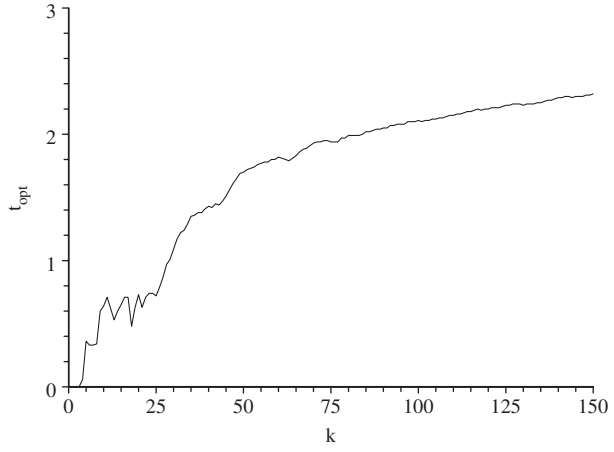

Fig. 3. Optimal value of t for each value of k.

For the Euclidean distance metric and $t = 0$, optimal values of k were quite different for different optimization criterion. In particular, for the RMSE criterion, $k = 25$ was optimal, whereas minimization of $\widehat{\mathrm{Var}}(\hat{\mu})$ where $\hat{\mu}$ is mean volume for a 4-km × 4-km AOI required $k > 100$ (Fig. 4).

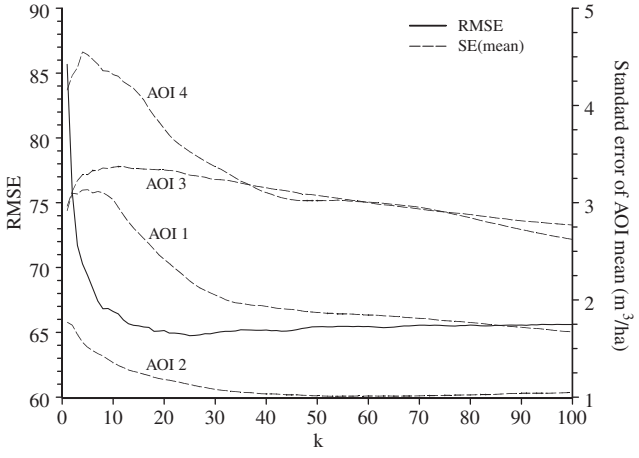

Fig. 4. Competing criteria relative to selection of the value of k.

5.3. Diagnostics

For k-NN predictions obtained using the Euclidean distance metric, the best combination of six feature variables with $k = 25$ (Table 1) and $t = 0$, the graph of group observation means versus group prediction means revealed no serious lack of fit (Fig. 5).

The residuals analyses revealed some very large standardized residuals; further, some of the reference set observations corresponding to large standardized residuals were used a large number of times as neighbors (Fig. 6) indicating their potential as influential observations. However, an assessment of the effects on RMSE of deleting reference set pixels with large standardized residuals, particularly those used a large number of times, indicated none was particularly influential. In fact, as evidenced by ratios of RMSEs for reference sets with potential influential observations deleted and RMSE for the entire reference set all being greater than 1 (Fig. 7), RMSEs actually increased when potential influential reference set observations were deleted from the reference set.

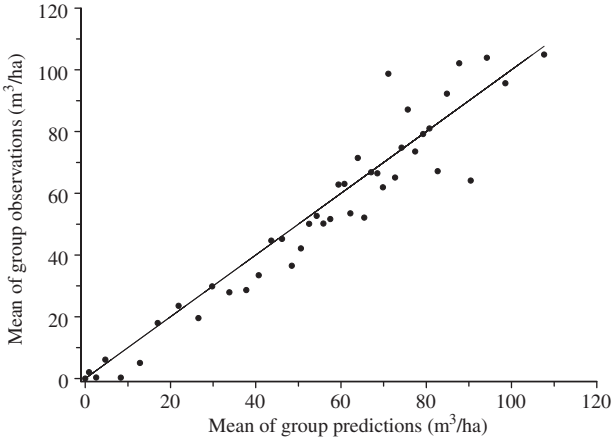

Fig. 5. Quality of fit.

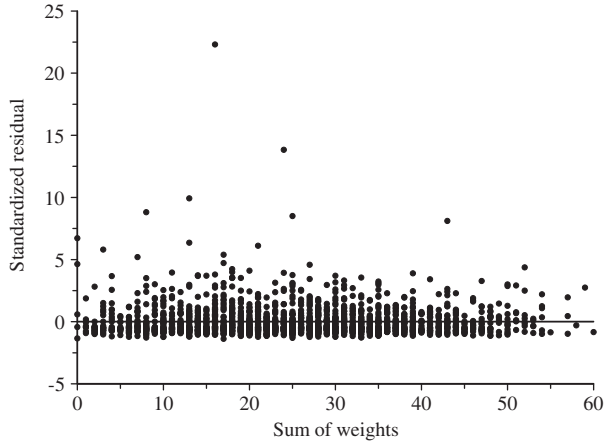

Fig. 6. Identifying potential outliers and influential reference set observations.

5.4. Inference

Differences between variance estimates obtained using the parametric and bootstrap approaches were small with proportional differences less in absolute value than 0.01 (Table 2). In addition, differences between estimates of bootstrap and parametric means were also small, indicating negligible bias in the bootstrap estimator of the mean. However, this bias assessment pertains only to the sampling aspects of the estimator, not to bias as a result of lack of fit which must be assessed separately (Fig. 5). These results validate the assumptions underlying the parametric approach but also suggest that the bootstrap approach merits consideration as an alternative to the more complex and computationally intensive parametric approach. Of particular note, the bootstrap approach should be expected to produce valid variance estimates for the MSN and GNN approaches for which $k = 1$, whereas estimates, $\hat{\sigma}_i^2$ from Eq. (6), may be questionable when using the parametric approach with small k. However, additional investigations should be conducted for small values of k. In addition, additional investigations are necessary when cluster sampling is used, a common feature of forest inventory programs.

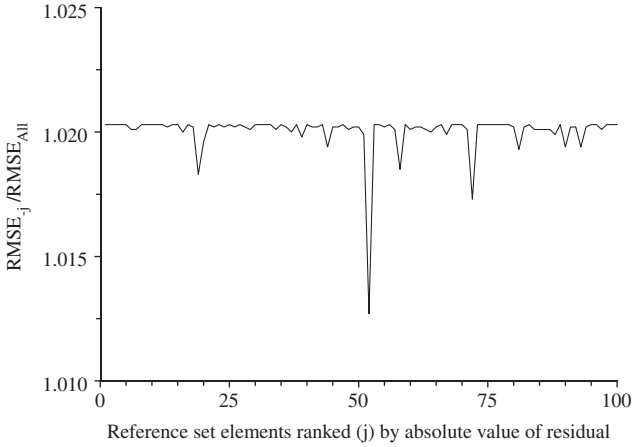

Fig. 7. Effects of potential influential reference set observations on RMSE; note all ratios are greater than 1.

Table 2
Parametric and bootstrap comparison.

Area of interest	Parametric		Bootstrap	
	$\hat{\mu}$	$\hat{\mathrm{Var}}(\hat{\mu})$	$\hat{\mu}_{boot}$	$\hat{\mathrm{Var}}_{boot}(\hat{\mu})$
1	64.75	3.65	64.76	3.67
2	36.14	1.97	36.07	1.96
3	49.29	3.18	49.45	3.05
4	81.52	5.08	81.40	5.02

6. Conclusions

Two primary conclusions may be drawn from the study. First, nearest neighbors techniques are a valid approach to estimation for small areas that lack sufficient numbers of inventory plots for traditional probability-based approaches including model-assisted methods. Comparisons of observations and nearest neighbor predictions indicated no lack of fit which, in turn, suggests the nearest neighbor estimator of the AOI volume means is unbiased or nearly unbiased. Further, the bootstrap variance estimator is a less complex and potentially less computationally intensive approach than the parametric approach. Regardless of the variance estimation approach, inferences in the form of confidence intervals for mean AOI volume are feasible. The second conclusion is that when using a combination of inventory plot and Landsat image data, the simple approach to k-NN prediction based on the Euclidean distance metric and equal neighbor weighting produced results that were comparable to results obtained using more complex and

computationally intensive optimization approaches. However, additional investigations should be conducted to determine if this result can be generalized to the multivariate case, particularly with larger numbers of feature variables.

References

Agarwal, A.K. (Ed.), 1977. Machine Recognition of Patterns. IEEE Press, New York.

Anderson-Sprecher, R., 1994. Model comparisons and R^2. The American Statistician 48 (2), 113–117.

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.

Belsey, D.A., Kuh, E., Welsch, R.E., 1980. Regression Diagnostics: Identifying Influential Data and Collinearity. Wiley, New York.

Breidenback, J., Northdurft, A., Kändler, G., 2010. Comparison of nearest neighbor approaches for small area estimation of tree species-specific forest inventory attributes in central Europe using airborne laser scanner data. European Journal of Forest Research 129, 833–846.

Chatterjee, S., Yilmaz, M., 1992. A review of regression diagnostics for behavioral research. Applied Psychological Measurement 16 (3), 209–227.

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.

Cover, T., Hart, P.E., 1967. Nearest neighbor pattern classification. IEEE Transactions on Information Theory IT-3, 21–27.

das Gupta, S., 1973. Theories and methods in classification: a review. In: Cacoullos, T. (Ed.), Discriminant Analysis and Applications. Academic Press, New York.

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

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.

Efron, B., Tibshirani, R., 1994. An introduction to the bootstrap. Chapman and Hall/CRC, Boca Raton, FL.

Eskelson, B.N.I., Temesgen, H., LeMay, V., Barrett, T.N., Crookston, N.L., Hudak, A.T., 2009. The role of nearest neighbor methods in imputing missing data in forest inventory and monitoring databases. Scandinavian Journal of Forest Research 24 (3), 235–246.

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), 130–135.

Fix, E., Hodges, J.L. 1951. Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. Report Number 4, Project 21-49-004. USAF School of Aviation Medicine. Randolph Field, Texas.

Fix, E., Hodges, J.L., 1989. Discriminatory analysis – nonparametric discrimination: consistency properties. International Statistical Review 57, 238–247.

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

Friedman, J.H. 1994. Flexible Metric Nearest Neighbor Classification. Technical Report. Department of Statistics, Stanford University. Stanford, California. 32p.

Fuchs, H., Magdon, P., Klein, C., Heiner, F., 2009. Estimating aboveground carbon in a catchment of the Siberian forest tundra: combining satellite imagery and field inventory. Remote Sensing of Environment 113, 518–531.

Gjertsen, A., 2007. Accuracy of forest mapping based on Landsat TM data and a kNN based method. Remote Sensing of Environment 110, 420–430.

Hastie, T., Tibshirani, R., 1996. Discriminant adaptive nearest neighbor classification. IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (6), 607–616.

Holland, J.H., 1975. Adaptation in natural and artificial systems. University of Michigan Press, Ann Arbor, Michigan.

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

Kajisa, T., Murakami, T., Mizoue, N., Kitahara, F., Yoshida, S., 2008. Estimation of stand volumes using the k-nearest neighbors method in Kyushu, Japan. Journal of Forest Research 13, 249–254.

Katila, M., Tomppo, E., 2005. Empirical errors of small area estimates from the multisource National Forest Inventory in Eastern Finland. Silva Fennica 40, 729–742.

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. In Proceedings of the Symposium on Machine Processing of Remotely Sensed Data. Purdue University, West Lafayette, IN, pp. 41–51.

Kendall, M.G., Buckland, W.R., 1982. A Dictionary of Statistical Terms, 4th ed. Longman Group, London.

Kim, J.S., Kim, Y.H., Kim, S.H., Jeong, J.H., Shin, M.Y., 2011. Comparison of the k-nearest neighbor technique with geographical calibration for estimating forest growing stock volume. Canadian Journal of Forest Research 41 (1), 73–82.

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

Kruskal, J.B., 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29 (1), 1–27.

Koukal, T., Suppan, F., Schneider, W., 2007. The impact of radiometric calibration on the accuracy of kNN-predictions of forest attributes. Remote Sensing of Environment 110, 431–437.

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 (2), 109–119.

LeMay, V.M., Maedel, J., Coops, N.C., 2008. Estimating stand structural details using nearest neighbor analyses to link ground data, forest cover maps, and Landsat imagery. Remote Sensing of Environment 112, 2578–2591.

McInerney, D.O., Nieuwenhuis, M., 2009. A comparative analysis of kNN and decision tree methods for the Irish National Forest Inventory. International Journal of Remote Sensing 30 (19), 4937–4955.

McRoberts, R.E., 2008. Using satellite imagery and the k-nearest neighbors technique as a bridge between strategic and management forest inventories. Remote Sensing of Environment 112, 2212–2221.

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., 2010a. 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., 2010b. 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., Nelson, M.D., Wendt, D.G., 2002. Stratified estimation of forest area using satellite imagery, inventory data, and the k-nearest neighbors technique. Remote Sensing of Environment 82, 457–468.

McRoberts, R.E., Tomppo, E.O., Finley, A.O., Heikkinen, J., 2007. Estimating areal means and variances 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.

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., 2010a. 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., 2010b. A model-assisted kNN approach to remove extrapolation bias. Scandinavian Journal of Forest Research 25, 174–184.

Maltamo, M., Eerikäinen, K., 2001. The most similar neighbour reference in the yield prediction of Pinus kesiya stands in Zambia. Silva Fennica 35 (4), 437–451.

Maselli, F., Chirici, G., Bottai, L., Corona, P., Marchetti, M., 2005. Estimation of Mediterranean forest attributes by the application of k-NN procedure to multitemporal Landsat ETM+ images. International Journal of Remote Sensing 26 (17), 3781–3796.

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

Nilsson, M., Holm, S., Wallerman, J., Engberg, J. 2005. Improved forest statistics from the Swedish National Forest Inventory by combining field data and optical satellite data using post-stratification. In: Olsson, H. (Ed.), Proceedings of ForestSAT 2005, 31 May–03 June 2005. Swedish Forest Agency, Borås, Sweden, pp. 22–26.

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.

Ohmann, J.L., Gregory, M.J., Henderson, E.B., Roberts, H.M., 2011. Mapping gradients of community composition with nearest-neighbour imputation: extending plot data for landscape analysis. Journal of Vegetation Science 22, 660–676.

Pierce Jr., B., Ohmann, J.L., Wimberly, M.C., Gregory, M.J., Fried, J.S., 2009. Mapping wildland fuels and forest structure for land management: a comparison of nearest neighbor imputation and other methods. Canadian Journal of Forest Research 39, 1901–1916.

Rajaniemi, S., Tomppo, E., Ruokolainen, K., Tuomisto, H., 2005. Estimating and mapping pteridophyte and Melastomataceae species richness in western Amazonian rainforests. International Journal of Remote Sensing 26 (3), 475–493.

Reese, H., Nilsson, M., Granqvist Pahlén, T., Hagner, O., Joyce, S., Tingelöf, U., 2003. Countrywide estimates of forest variables using satellite data and field data from the National Forest Inventory. Ambio 32, 542–548.

Ricci, F., Avesani, P., 1996. Data compression and local metrics for nearest neighbor classification. IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (4), 380–384.

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

Salovaara, K.J., Thessler, S., Malik, R.N., Tuomisto, H., 2005. Classification of Amazonian primary rain forest vegetation Landsat ETM+ satellite imagery. Remote Sensing of Environment 97, 39–51.

Silverman, B.W., Jones, M.C., 1989. E. Fix and J.L. Hodges (1951): An important contribution to nonparametric discriminant analysis and density estimation. International Statistical Review 57, 233–247.

Thessler, S., Sesnie, S., Bendaña, Z.S.R., Ruokolainen, K., Tomppo, E., Finegan, B., 2008. Using k-nn and discriminant analyses to classify rain forest types in a Landsat TM image over northern Costa Rica. Remote Sensing of Environment 112 (5), 2485–2494.

Temesgen, H., Barrett, T.M., Latta, G., 2008. Estimating cavity tree abundance using nearest neighbor imputation methods for western Oregon and Washington forests. Silva Fennica 42 (3), 337–354.

Temesgen, H., LeMay, V.M., Froese, K.L., Marshall, P.L., 2003. Imputing tree-lists from aerial attributes for complex stands of south-eastern British Columbia. Forest Ecology and Management 177, 277–285.

Tomppo, E. 1991. Satellite image-based national forest inventory of Finland. In: Proceedings of the Symposium on Global and Environmental Monitoring, Techniques and Impacts. 17–21 September 1990. Victoria, British Columbia, Canada. International Archives of Photogrammetry and Remote Sensing 28, 419–424.

Tomppo, E. 1996. Multi-source national forest inventory of Finland. In: Päivinen, R., Vanclay, J., Miina, S. (Eds.), New Thrusts in Forest Inventory, Proceedings of the Subject Group S4.02-00, Forest Resource Inventory and Monitoring, and Subject Group S4.12-00, Remote Sensing Technology, vol. I, IUFRO XX World Congress, Tampere, Finland, 6–12 August 1995. EFI, EFI Proceedings 7, 27–41.

Tomppo, E., 2006. The Finnish National Forest inventory. In: Kangas, A., Maltamo, M. (Eds.), Forest Inventory: Methodology and Applications. Springer, Dordrecht, The Netherlands.

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

Tomppo, E., Goulding, C., Katila, M., 1999. Adapting Finnish multi-source forest inventory techniques to the New Zealand preharvest inventory. Scandinavian Journal of Forest Research 14, 182–192.

Tomppo, E., Haakana, M., Katila, M., Peräsaari, J., 2008. Multi-source National Forest Inventory. Springer, 373 p.

Tomppo, E., Korhonen, K.T., Heikkinen, J., Hannu, Y.-L., 2001. Multi-source inventory of forests of the Hebei Forestry Bureau, Heilongjian, China. Silva Fennica 35 (3), 309–328.

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.

Xu, X., Du, H., Zhou, G. 2011. Estimating Moso bamboo forest attributes using the k nearest neighbors technique and satellite imagery. Photogrammetric Engineering and Remote Sensing.