Remote Sensing of Environment 113 (2009) 476–488

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Model-based mean square error estimators for *k*-nearest neighbour predictions and applications using remotely sensed data for forest inventories

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

[a] Natural Resources Canada, Canadian Forest Service, 506 West Burnside Road, Victoria BC, Canada V8Z 1M5
[b] USDA Forest Service, Northern Research Station, St. Paul, Minnesota, MN 55108 USA
[c] The Finnish Forest Research Institute, Helsinki, FIN-00170, Finland

ARTICLE INFO

Article history:
Received 27 November 2007
Received in revised form 10 April 2008
Accepted 12 April 2008

Keywords:
Small area estimation
Spatial prediction
Non-parametric
Single index model
Variance function
Spatial correlation function
Forest inventory

ABSTRACT

New model-based estimators of the uncertainty of pixel-level and areal *k*-nearest neighbour (k_{nn}) predictions of attribute *Y* from remotely-sensed ancillary data **X** are presented. Non-parametric functions predict *Y* from scalar 'Single Index Model' transformations of **X**. Variance functions generated estimates of the variance of *Y*. Three case studies, with data from the Forest Inventory and Analysis program of the U.S. Forest Service, the Finnish National Forest Inventory, and Landsat ETM+ ancillary data, demonstrate applications of the proposed estimators. Nearly unbiased k_{nn} predictions of three forest attributes were obtained. Estimates of mean square error indicate that k_{nn} is an attractive technique for integrating remotely-sensed and ground data for the provision of forest attribute maps and areal predictions.

1. Introduction

For a small area of interest (AOI, a finite set of units for which an estimate is desired), a direct estimate (prediction) of the total of an attribute from a sample within the AOI is often discouraged by a small sample size (Hansen et al., 1983; Rao, 2003). The situation can be improved if one has ancillary data for each unit in the AOI that are correlated with the attribute(s) of interest. A clever use of this ancillary information for stratification, regression, or calibration can reduce the probability-based mean square error (MSE) of desired estimates (D'Orazio et al., 2006; Wu, 2003). Model-based or model-assisted approaches also exploit the information about *Y* contained in (**X**) and have become popular in practice (Flores & Martínez, 2000; Singh et al., 2002; Zhang & Chambers, 2004). In forest inventories, several (e.g., *p*) attributes – often a mix of correlated continuous, ordinal, and categorical attributes – are of interest which makes the estimation problem multivariate. Furthermore, maps showing the unit-level spatial distribution of predicted values in the AOI – consistent with the estimated totals – are usually also desired. An optimized attribute-by-attribute estimation procedure would quickly become an onerous burden inasmuch as the relationship between the ancillary **X** and *Y* is attribute dependent, often complex (non-linear), and usually with a heteroscedastic variance structure and spatial correlation in both **X** and *Y*. As well, unit-level predictions should, as much as possible, be consistent with the observed variation, range, and interdependencies of *Y*.

A popular choice in forest inventory is the non-parametric and distribution free *k*-nearest neighbour method, or k_{nn} (Alt, 2001). In k_{nn} the predictions of attribute values for a unit (target) in the AOI are linear combinations of attribute-values in a set of *k* units selected from a so-called reference set of units with known values of *Y*. The choice of these units is determined exclusively by a distance-metric defined on the ancillary variable space. The *k* reference units with the smallest distances to the target unit in the ancillary space are selected. It is the simultaneous prediction of all attributes of interest from a linear combination of the attributes of the selected units that distinguish the k_{nn} procedure from more conventional prediction approaches. In k_{nn} a univariate prediction is not materially different from a multivariate prediction because i) the same *k* reference units are used for each attribute of interest, and ii) the prediction is a linear combination of the attribute values of the reference units. It is a non-parametric estimator since predictions can be made without estimating any parameters. It is also a distribution free procedure because predictions can be made without the need for any distributional assumptions. As we shall see, however, distributional assumptions are needed for estimating the uncertainty of k_{nn} predictions. Thus ease of multivariate predictions, implementation (non-parametric, distribution free), and flexibility are the most compelling attractions of k_{nn}. Consistency of multivariate

* Corresponding author.
E-mail address: steen.magnussen@nrcan.gc.ca (S. Magnussen).

0034-4257/$ – see front matter.
doi:10.1016/j.rse.2008.04.018

associations and range of predicted values are also regarded as attractive features, although this is strictly possible only for $k=1$ (Moeur et al., 1995). Today k_{nn} is widely used for small AOI forest inventory estimation problems (Collins et al., 2004; Franco-Lopez et al., 2001; LeMay & Temesgen, 2005; McRoberts et al., 2002; Reese et al., 2002; Tomppo, 2006). We consider k_{nn} predictions as model-based (McRoberts et al., 2007; Valliant et al., 2000 ch. 3.7), which means that the analyst regards unit-level y-values as random realizations from a super-population and $\mathbf{X}$ as fixed by design. The super-population has a finite set of N units with fixed values of $\mathbf{X}$ (i.e. $\mathbf{x}$) and Y (i.e. y) generated by a stochastic process described by the model. Hence, an infinite number of ys (with probabilities defined by model assumptions) can be generated for a given unit in the AOI. Given an observed ancillary value (i.e. $\mathbf{x}$), the realizations y follow an assumed distribution. A common distributional assumption is the Gaussian; fully specified by its mean $\mu_y(\mathbf{x})$ and variance $\Gamma_y(\mathbf{x})$. In a k_{nn} context a prediction of Y is desired for each unit in a target set. Predictions can be of the expected mean or the random realization, but in the context of k_{nn} prediction we are only concerned with the prediction of realizations which are random variables. Examples of both prediction of unit-level realizations and areal means can be found in McRoberts et al. (2007). Reference units are selected on the basis of a similarity index defined on the feature space of the ancillary variables. In applications, where the total of unit-level ys in an AOI is to be predicted, a model-based inference must consider the effects of spatial covariance of deviations of Y from their expectations. With models, commonly derived from a single reference set of data, the risk of model-bias cannot be ignored.

A recognized drawback of k_{nn} has been the lack of an unbiased and consistent estimator of the uncertainty in unit-level and areal k_{nn} predictions for an AOI. Until recently, the root mean square error (RMSE) estimated for a single prediction in a leave-one-out cross-validation procedure has served as the only measure of uncertainty. Shortcomings relate to i) differences in the data domain of the reference and target units, ii) omission of covariances between predicted values, iii) omission of potential spatial correlations of Y given $\mathbf{X}$, and iv) violation of an implicit assumption of variance homogeneity and independence of Y. A covariance between predictions will add in a non-trivial manner to the uncertainty of a predicted total for an AOI (Koistinen et al., 2008; McRoberts et al., 2007). Further, an RMSE from a leave-one-out crossvalidation procedure has a built-in positive bias by virtue of dropping the nearest neighbour from the reference set (Chen & Shao, 2001).

In a forest inventory context, the first to propose an estimator of the uncertainty of a k_{nn} prediction were Moeur et al. (1995). In an application of a most-similar-neighbour procedure they suggested a linear relationship between the squared prediction error and the Mahalanobis distance in the ancillary space. No details were given, however. Despite an intuitive appeal, our simulations (not shown) with a univariate Gaussian attribute and standardized multivariate Gaussian ancillary variables suggest that both the Mahalanobis as well as other common distance metrics (Euclidean, Manhattan, Chebyshev) are poor predictors of the squared prediction error when the dimension of $\mathbf{X}$ is greater than or equal to 3 and when the correlation between Y and the ancillary variables is weak.

Chen and Shao (2001) proposed a jackknifed variance estimator. They observed that a naive jackknifed variance estimator seriously underestimates the variance by treating predictions as if they were observations. Conversely, an iteration of the predictions using previously predicted values led to inflated estimates of the error variance. As a compromise, they proposed a modified jackknifed procedure with partial replicate predictions and partial adjustments of pseudo-replications. They proved that their procedure is asymptotically unbiased and consistent. Their variance estimator is limited to predictions using $k=1$ and cannot immediately be extended to $k>1$. The performance of this variance estimator is expected to decline rapidly as the ratio of the number of target units to the number of reference units increases. In small area estimation, this ratio is usually large.

Kim and Tomppo (2006) proposed a variogram model-based estimation of the error variance of unit-level k_{nn} predictions of volume per ha. A Matérn class variogram model was estimated from spatially de-trended data and used to predict the covariance between residuals (predicted–observed) as a function of their distance in a reduced ancillary space spanned by the first two principal components of the covariance matrix of six ancillary Landsat 5TM variables. Variograms were largely dominated by 'nugget' effects, which made the 'kriging' predictions and the prediction variance approximately equal to, respectively, the mean and variance of the reference units. In a validation trial the estimate of the RMSE compared reasonably well with empirical estimates obtained in a leave-one-out crossvalidation. Common problems of i) de-trending residuals, ii) heterogenous variance, and iii) estimating the 'smoothness parameter' ν in Matérn's variogram model were duly recognized (see also Glatzer & Muller, 2004; Marchant & Lark, 2004; Zhao & Wall, 2004). A stratification of the land-base with a separate variogram model for each stratum seems imperative in practical applications. Stratification increases the number of reference units needed to estimate the variogram models. Kim and Tomppo's (2006) cubic transformation of attribute values may have generated an impression of an improved error-estimation. However, a back-transformation to the scale of interest would negate this apparent improvement. Furthermore, the variance estimator does not seem to account fully for a possible correlation between attribute values of the k selected reference units. This is a correlation that can arise from a feature-space distance ranking.

McRoberts et al. (2007) proposed a model-based variance estimator for both unit-level and areal predictions. The variance of random realizations is taken as the sum of squared deviations of the k selected reference attribute values from the k_{nn} prediction divided by the effective degrees of freedom calculated as a function of the correlation between the reference units (Cliff & Ord, 1981). A variance estimator for the average of the attribute in the AOI is consequently obtained by summing the estimated variances and covariances of all units in the AOI (Royall, 1988). McRoberts et al. (2007) estimators rely on two key assumptions regarding the attribute values of target unit and the k reference units: i) equality of expected values, and ii) equality of the variances. McRoberts et al. chose a simple approach to estimation in which all the nearest neighbours were weighted equally. As a result, their variance estimator does not fully take into account that a distance-ordering of reference units can introduce a distinct covariance structure (David & Mishriky, 1968). Finally, by virtue of the assumption that the expected values of the k_{nn} reference units are approximately equal to the expected value of the target unit, bias is not considered.

An extension of the univariate variance estimation approach in McRoberts et al. (2007) to a multivariate Y requires an estimation of the among-attribute covariance of reference units and the effective degrees of freedom for the variance estimates to be derived as the trace of the 'Hat matrix' linking a k_{nn} prediction to the reference attribute values (Ruppert et al., 2003).

A new model-based estimator of the uncertainty of k_{nn} predictions is proposed in this study. It distinguishes itself from the above studies through: i) an estimation of the expected value of Y by a univariate non-parametric regression (g) using a scalar X^* obtained via a single index model (SIM) transformation of $\mathbf{X}$ as a predictor of Y (cf. Appendix A.1); ii) an estimator of variance (Γ) for predicting the expected variance of Y given X^*; iii) considering the correlation introduced by the distance ordering and weighting of the k reference Y-values used in a k_{nn} prediction; and iv) a provision of an estimate of bias of a k_{nn} prediction of an expected Y-value. An extension to multivariate estimation and predictions preserves the above distinction.

2. Materials and methods

2.1. Notation

A variable typed in upper case refers to the attribute while a lower case typing indicates the realized (observed) value of the attribute. Multivariate variables are set in bold to indicate a vector or a matrix. Univariate variables are set in normal type. Greek letters are reserved for model parameters. We use a ~ (tilde) to denote a prediction of a random variable, and a ^ (hat/caret) to indicate a probability-based estimate of a finite population statistic of interest, like the mean ($\bar{y}$) and total (T_y) of Y and associated estimates of sampling errors.

In this paper the variable (attribute) of interest is considered to be univariate. An extension to the multivariate case with p attributes of interest is straightforward as it raises no new issues except for a necessary switch to matrix algebra (Searle, 1982). Since the reference units are selected on the basis of the ancillary attribute values the same reference units are selected for each attribute of interest. Our case studies include two and three attributes but since we make no use of the covariance between predicted attribute values a univariate approach suffices. A univariate presentation is also easier to follow. For those who need to estimate multivariate functions of predicted values, a multivariate approach will be needed. Key aspects of the multivariate estimation procedure are in appendices.

Our proposed attribute-specific transformation of the q ancillary $\mathbf{X}$ via a SIM procedure to a scalar X^* is not prerequisite but only illustrates a potential improvement of the efficiency of the k_{nn} approach to prediction. To those who include a SIM transformation step, their ancillary $\mathbf{X}$ is $(X_1^*,\ldots, X_p^*)$. Details of SIM are in Appendix A.1.

In this study the terms unit and pixel are synonymous. We use both and let the context determine usage.

2.2. Target and reference set

Let N denote the number of units in an area of interest (AOI) for which we want a prediction of Y, and let $\mathbf{X}$ be a row vector of q ancillary variables carrying information about Y. We refer to these N units as the target set for k_{nn} predictions. A reference set of n units representative of the units in the AOI with Y and $\mathbf{X}$ recorded for each unit j ($j=1,\ldots,n$) is available as predictors. Some, none, or all of the reference units may be part of the target set (AOI). Ancillary information is known for all units in the target set and the information is viewed as fixed (non-random).

2.3. Population model

For a given unit (e.g., i) with ancillary vector $\mathbf{x}_i$ the attribute value of interest is assumed to be a random realization y_i from a super-population with a fixed (unobservable) super-population mean $\mu_y(\mathbf{x}_i)$, a variance $\Gamma_y(\mathbf{x}_i)$ that depends on $\mathbf{x}_i$, and spatial correlation of realizations. To simplify notation, we shall use $\mu_i \equiv \mu_y(\mathbf{x}_i)$ and $\Gamma_i \equiv \Gamma_y(\mathbf{x}_i)$ whenever the implicit dependence on $\mathbf{x}_i$ is clear from the context. Attribute values associated with the target and the reference sets are assumed to be generated from the same super-population. Specifically, we have

$$y_i = \mu_i + \delta_i \tag{1}$$

where δ_i is a random deviation from the expected mean with expectation $E(\delta_i)=0$ and $\text{var}(\delta_i)=\Gamma_i$. A unit-level expected value, μ_i, of y_i is with respect to all possible unit-specific realizations from the super-population.

2.4. Unit level k_{nn} predictions

Our objective is a k_{nn} prediction of each unit level realization y_i of all target units (N) in the AOI. From these predictions, a prediction of the total T_y or a unit level average in the AOI is obtained. Unit level predictions are also used for mapping purposes. A measure of uncertainty is required for both unit level and AOI predictions of the total ($\tilde{T}_y$) which is simply the sum of N k_{nn} predictions. The unit average is $\mu_y = T_y \times N^{-1}$ and its k_{nn} estimator is $\tilde{\mu}_y = \tilde{T}_y \times N^{-1}$. The mean square error (variance plus bias squared) or MSE serves as the measure of uncertainty.

A k_{nn} prediction of the random realization y_i for a target unit $i=1,\ldots,N$ was computed from the reference units as a weighted mean of the attribute values associated with the k reference units that are nearest in terms of the ancillary values ($\mathbf{X}$) to target unit i

$$\tilde{y}_i = \sum_{u=1}^{k} w_{u(i)} y_{u(i)} \text{ with } \sum_{u=1}^{k} w_{u(i)} \equiv 1, \quad i = 1,\ldots,N \tag{2}$$

where $y_{u(i)}$ denotes the value of a variable observed for a reference unit, $u(i)$, that is the uth nearest neighbour to the target unit i, taken from the reference set of n units, and $w_{u(i)}$, the weight given to the uth nearest reference unit (Aha, 1997). In this study weights are inversely proportional to the Euclidian distance in the ancillary space ($\mathbf{X}$ distance) between the target unit i and the reference unit $u(i)$, and they are scaled to sum to one. An important consequence – of this weighting scheme – is that a k_{nn} prediction for a target unit becomes equal to the observed value if the unit is also a reference unit. This follows from the zero $\mathbf{X}$ distance between a target and a reference unit when they are identical. A zero distance means that the inverse to the distance is infinite; hence the weight given to a reference unit with zero distance to the target will be arbitrarily close to 1. Should there be more than one reference unit with an ancillary vector identical to that of the target unit, then one would have to perturb them by adding a small amount of white noise to resolve a mathematical impasse.

In a logical extension of the super-population model in (1) we can write $\tilde{y}_i$ as the sum of a super-population prediction $\tilde{\mu}_i$ and a residual term $\tilde{\varepsilon}_i$ with mean zero and variance $\tilde{\omega}_i$

$$\tilde{y}_i = \tilde{\mu}_i + \tilde{\varepsilon}_i \tag{3}$$

Given that a k_{nn} prediction is a linear combination of k realized random variables from the reference set, a consistent estimator of the expected value of a k_{nn} prediction is $\tilde{\mu}_i = \sum_{u(i)=1}^{k} w_{u(i)} \mu_{u(i)}$ which leads to $\tilde{\varepsilon}_i = \sum_{u=1}^{k} w_{u(i)} \delta_{u(i)}$, and the following estimator of the variance of $\tilde{\varepsilon}_i$: $\tilde{\omega}_i = \sum_{u=1}^{k} \sum_{v=1}^{k} w_{u(i)} w_{v(i)} \text{cov}(\delta_{u(i)}, \delta_{v(i)})$.

2.5. MSE of a k_{nn} unit level prediction

The uncertainty of a unit level k_{nn} prediction is quantified by the MSE

$$\begin{aligned} \text{MSE}(\tilde{y}_i) &= E(\tilde{y}_i - y_i)^2 = E[(\tilde{y}_i - \tilde{\mu}_i) + (\tilde{\mu}_i - \mu_i) + (\mu_i - y_i)]^2 \\ &= E(\tilde{y}_i - \tilde{\mu}_i)^2 + E(\mu_i - y_i)^2 + E(\tilde{\mu}_i - \mu_i)^2 + \\ &\quad 2E[(\tilde{y}_i - \tilde{\mu}_i)(\tilde{\mu}_i - \mu_i)] + 2E[(\tilde{y}_i - \tilde{\mu}_i)(\mu_i - y_i)] + 2E[(\tilde{\mu}_i - \mu_i)(\mu_i - y_i)] \\ &= \text{var}(\tilde{y}_i - \tilde{\mu}_i) + \text{var}(y_i - \mu_i) + (\tilde{\mu}_i - \mu_i)^2 - 2\text{cov}(\tilde{y}_i - \tilde{\mu}_i, y_i - \mu_i) \\ &= \tilde{\omega}_i + \Gamma_i - 2\tilde{\omega}_i^{0.5} \Gamma_i^{0.5} \eta(\tilde{\varepsilon}_i, \delta_i) + (\tilde{\mu}_i - \mu_i)^2 \end{aligned} \tag{4}$$

where η is the correlation coefficient between $\tilde{\varepsilon}_i = \tilde{y}_i - \tilde{\mu}_i$, and $\delta_i = y_i - \mu_i$, and where expectations are taken over all possible realizations of y_i. Going from the second to the third line of Eq. (4) requires the assumption that the bias term $(\tilde{\mu}_i - \mu_i)$ is independent of the residual ε_i and the deviation δ_i (i.e. $E(\tilde{y}_i - \tilde{\mu}_i)(\tilde{\mu}_i - \mu_i) = 0$ and $E(y_i - \mu_i)(\tilde{\mu}_i - \mu_i) = 0$). The last line in Eq. (4) follows from Eqs. (1) and (3) and previous definitions of Γ_i and $\tilde{\omega}_i$, and the fact (Cliff & Ord, 1981) that $\text{cov}(\tilde{y}_i - \tilde{\mu}_i, y_i - \mu_i) = \text{var}(\tilde{y}_i - \tilde{\mu}_i)^{0.5} \text{var}(y_i - \mu_i)^{0.5} Corr(\tilde{y}_i - \tilde{\mu}_i, y_i - \mu_i)$.

In practice, to obtain an estimate of $\text{MSE}(\tilde{y}_i)$ one must substitute predictions of the unknowns in Eq. (4). Appendices A.1–A.5 outline procedures for obtaining these (model-based) predictions from the reference set. Mean square error results are mostly scaled to relative root mean square errors (RRMSE) which is the root mean square error

$(\mathrm{RMSE} = \sqrt{\mathrm{MSE}})$ divided by the estimate (prediction, or actual value) to which the MSE applies. An empirical estimate of $\mathrm{MSE}(\tilde{y}_i)$ was obtained from the reference set via a leave-one-out crossvalidation procedure on the reference set. It is denoted as $\widehat{\mathrm{MSE}}_{\mathrm{CV}}(\tilde{y}_i)$.

2.6. A k_{nn} estimator of T_y

The k_{nn} estimator of T_y the total of Y in the AOI is

$$\tilde{T}_y = \sum_{i=1}^{N} \tilde{y}_i \tag{5}$$

where, as noted before, the k_{nn} predictions for the reference units in an AOI will, given our weighting scheme, always be very close to their observed values. Accordingly, the k_{nn} estimator of the unit average of is $\tilde{\mu}_y = \tilde{T}_y / N$.

2.7. An estimator of the MSE of $\tilde{T}_y$

A model-based estimation of the MSE of $\tilde{T}_y$ in Eq. (5) leads to:

$$\begin{aligned}\mathrm{MSE}\left(\tilde{T}_y\right) &= E\left(\tilde{T}_y - T_y\right)^2 = E\left\{\sum_{i=1}^{N} \tilde{y}_i - \sum_{i=1}^{N} y_i\right\}^2 = E\left\{\sum_{i=1}^{N} \left((\tilde{y}_i - \tilde{\mu}_i) + (\tilde{\mu}_i - \mu_i) + (\mu_i - y_i)\right)\right\}^2 \\ &= E\left\{\sum_{i=1}^{N}\sum_{i'=1}^{N} \left((\tilde{y}_i - \tilde{\mu}_i) + (\tilde{\mu}_i - \mu_i) + (\mu_i - y_i)\right)\left((\tilde{y}_{i'} - \tilde{\mu}_{i'}) + (\tilde{\mu}_{i'} - \mu_{i'}) + (\mu_{i'} - y_{i'})\right)\right\} \\ &= \sum_{i=1}^{N}\sum_{i'=1}^{N} \left\{\mathrm{cov}(\tilde{y}_i, \tilde{y}_{i'}) + \mathrm{cov}(y_i, y_{i'}) - 2\mathrm{cov}(\tilde{y}_i, y_{i'})\right\} + N\sum_{i=1}^{N} (\tilde{\mu}_i - \mu_i)^2.\end{aligned} \tag{6}$$

To get from the first to the second line of Eq. (6) we have made use of the relationship $(\sum_l z_l)^2 = \sum_l \sum_m z_l z_m$ which follows from a simple expansion and rearrangement of terms. One gets to the third line in Eq. (6) by multiplying out the terms and taking the expectation over all possible realizations, assuming that i) covariance between the bias of a k_{nn} prediction $(\tilde{\mu}_i - \mu_i)$ is independent of the stochastic deviation (δ_i) and the residual $(\tilde{\varepsilon}_i)$, and ii) that bias terms for distinct pairs of predictions (i,i') are independent; that is:

$$\begin{aligned}&\mathrm{cov}(\tilde{y}_i - \tilde{\mu}_i, \tilde{\mu}_{i'} - \mu_{i'}) \equiv 0, \\ &\mathrm{cov}(y_i - \mu_i, \tilde{\mu}_i - \mu_{i'}) \equiv 0, \text{ and} \\ &\mathrm{cov}(\tilde{\mu}_i - \mu_i, \tilde{\mu}_{i'} - \mu_{i'}) \equiv 0,\ i' \neq i\ .\end{aligned} \tag{7}$$

The first of the three assumptions can be rewritten into the covariance between two linear combinations of random deviations (δ) which should be zero by our super-population assumptions. The second can be rewritten into a covariance between one random deviation and a linear combination of different δs, also zero by definition. Finally, the third assumption was evaluated. We found that estimates of this covariance were relatively small and negative. Our RRMSE estimates which do not include this covariance are thus slightly too high (conservative). Results for basal area per ha (BA) suggest that our RRMSE have been inflated by less than 0.1% by dropping the covariance of bias terms.

Our estimator of the MSE in Eq. (6) becomes

$$\begin{aligned}\widehat{\mathrm{MSE}}\left(\tilde{T}_y\right) &= \hat{E}\left(\tilde{T}_y - T_y\right)^2 \\ &= \sum_{i=1}^{N}\sum_{i'=1}^{N} \left\{\widehat{\mathrm{cov}}(\tilde{y}_i, \tilde{y}_{i'}) + \widehat{\mathrm{cov}}(y_i, y_{i'}) - 2\widehat{\mathrm{cov}}(y_i, \tilde{y}_{i'})\right\} + N\sum_{i=1}^{N} (\tilde{\mu}_i - \hat{\mu}_i)^2\end{aligned} \tag{8}$$

where estimates $\hat{\mu}_i$ are obtained from $\hat{g}(x_i^*)$ as determined by the SIM model detailed in A.1.

Details on estimation procedures for the covariance terms in Eq. (8) are given in Appendices A.6–A.8. The last term in Eq. (8), i.e., the sum of the estimated squared bias of k_{nn} predictions of the mean i.e., $(\tilde{\mu}_i - \hat{\mu}_i)^2$, $i = 1,\ldots,N$ was computed directly from unit-level predictions and estimates of μ_i.

Estimating and summing the covariance terms in Eq. (8) can be prohibitively time-consuming when N or k or both are large. A fast yet accurate sample-based approximation to these sums is recommended when time to compute becomes an issue (McRoberts et al., 2007).

3. Case studies

Three forest inventory data sets (FI1, FI2, FI3) with multivariate attributes of interest and co-located information on a suite of ancillary remotely-sensed attributes were used to illustrate applications of the proposed estimators of the uncertainty in k_{nn} predictions. Each data set is a probability sample. Inventory estimates of a unit average of y $(\bar{y})$ are denoted as $\hat{\bar{y}}$ while an estimate of an areal total is $\hat{T}_y$. Estimators of MSE of probability-based inventory samples were computed as detailed in Cochran (1977) for single stage cluster-sampling.

3.1. FI1 and FI2

The two data sets were provided by the Forest Inventory and Analysis (FIA) program of the U.S. Forest Service. Both are from forested areas in Minnesota. A unit is a Landsat pixel approximately 30 m × 30 m in size. A total of 3124 (FI1) and 5195 (FI2) subplots (pixels) from 872 (FI1) and 1492 (FI2) inventory sampling locations provided the pixel-specific attribute and ancillary data. No additional data was used; the AOIs are therefore composed entirely of units with co-located FIA subplots. The FIA sampling frame (Bechthold & Patterson, 2005) is a network of hexagonal cells. Each hexagon is approximately 2402.6 ha with a distance between hexagon centers of approximately 6.08 km. One sample location with four subplots in a cluster is located at random within each hexagon. The cluster of four subplots at an inventory location is configured as a central subplot and three peripheral subplots with centers located at 36.58 m and azimuths of 0°, 120°, and 240° from the center of the central subplot. A subplot is circular with a radius of 7.31 m. Almost all subplots were either completely forested or completely without forest. Only forested subplots are included in this study which brought the average number of subplots per sample location down to 3.6 for FI1 and 3.5 for FI2. Ancillary attribute data came from Landsat 7 ETM+ scenes in path 26 and row 27 (FI1), and in path 27 and row 27 (FI2). Imagery for three dates representing early, peak, and late vegetation green-up were used. FI1 images dates were May 2003, June 2001, and November 1999. Corresponding dates for FI2 were April 2000, July 2001, and November 1999. Sub-plots were geo-located with GPS to an accuracy of 0.5 pixels and matched to pixel centers. Each sub-plot covers approximately 19% of a pixel.

Four ancillary variables were extracted from each scene: NDVI and tasselled cap transformations, brightness, greenness, and wetness (Kauth & Thomas, 1976). They were selected based on the strength of their association with the attributes of interest (McRoberts, 2006). Hence, a total of $q = 12$ ancillary covariates were included.

Attributes of interest were i) number of trees per ha (TPH), ii) basal area in square meter per ha (BA), and iii) volume in cubic meter per ha (VOL). Only trees with a diameter ≥12.7 cm at a reference height of 1.37 m aboveground were included in the reference estimates of TPH, BA, and VOL.

FI1 and FI2 data were split at random into a reference set (60%) and an AOI viz. target set (40%) for which pixel-level k_{nn} predictions are desired. The separation into a target and a reference set was done at the subplot level. Accordingly, the size of the reference set (n) was 1875 for FI1, and 3117 for FI2. Corresponding numbers for the target sets (N) were 1249 and 2078, respectively. Note that the artificial AOIs constructed for FI1 and FI2 do not constitute compact contiguous spatial areas. This is of consequence for the expected covariance between k_{nn} predictions. All models and parameters needed to estimate the uncertainty of either a pixel-level k_{nn} prediction or a sum of pixel-level

Table 1
FI1 inventory and k_{nn} predictions

	Reference set			Target set		
	BA $m^2 \times ha^{-1}$	VOL $m^3 \times ha^{-1}$	TPH ha^{-1}	BA $m^2 \times ha^{-1}$	VOL $m^3 \times ha^{-1}$	TPH ha^{-1}
Probability-based average $\left(\hat{\bar{y}}\right)$	18.9	81.2	1935	19.0	79.8	2034
k_{nn} prediction of average $(\tilde{\mu}_y)$	19.1 (0.7)	82.9 (2.2)	1913 (−1.1)	19.1 (0.5)	81.9 (2.7)	1930 (−5.1)
$\widehat{\mathrm{RRMSE}}\left(\hat{\bar{y}}\right)\%$[a]	1.9	2.8	2.8	2.2	3.2	3.2
$\widehat{\mathrm{RRMSE}}_{CV}(\tilde{y}_i)\times N^{-0.5}\%$	1.7	2.3	2.6	2.0	2.8	3.1
$\widehat{\mathrm{RRMSE}}(\tilde{y}_i)\times N^{-0.5}\%(cf.4)$	1.7	2.3	2.6	1.6	2.3	2.5
$\widehat{\mathrm{RRMSE}}\left(\tilde{T}_y\right)\%(cf.8)$				2.5	3.5	3.7

Number of pixels (subplots) is 1875 in the reference set (60%), and 1249 (40%) in the target set. All k_{nn} results are for $k=7$. Numbers in parentheses are the relative deviation between the k_{nn} prediction, and the estimate from a probability sample: $\left(\tilde{\mu}_y-\hat{\bar{y}}\right)\times\hat{\bar{y}}^{-1}\times 100$.
[a] Estimated as per Cochran (1977 p 250) for a single stage cluster sampling with unequal number of units (subplots) per sample-location (cluster).

predictions were obtained from the reference set either directly or through a k_{nn} leave-one-out crossvalidation procedure.

A reference pixel (subplot) to be used in a k_{nn} prediction for a target pixel, in a crossvalidation procedure, or in a model fitting procedure could not be from the same sample location as the pixel for which a prediction was desired. This exclusion was imposed to mimic a practical application of k_{nn}.

3.2. FI3

The AOI is three separate small (~100 ha) contiguous forested areas located in the eastern part of Central Finland (North Karelia and South Savo), approximately between latitudes 61°10′ N, 63°95′ N and longitudes 27°20′ E, 31°10′E. A population unit is a quarter of a Landsat 7 ETM+ image pixel, approximately 12.5 m×12.5 m in size. The total number of pixels (N) in the AOI is 17,620 (275.4 ha). A k_{nn} prediction of three forest inventory attributes – DBH (basal area weighted diameter of tree stems 1.3 m aboveground), BA (basal area in m^2 ha^{-1} of tree stems 1.3 m aboveground), and VOL (stem volume in m^3 ha^{-1}) – is desired for each (target) pixel classified as 'forest' or 'other wooded land' by a national land use classification system. All reference pixels (see below) were located outside the AOI. Predictions of the three attributes are done separately for two soil strata (mineral and peatland) as this stratification improves the accuracy of k_{nn} predictions (Katila & Tomppo, 2002; Tomppo & Halme, 2004). The breakdown of target pixels by soil type is 14,644 (83%) on mineral soils and 2976 (17%) on peatland. The pixel-level ancillary variables for both the target and the reference set are in the form of nine Landsat 7 ETM+ bands ($q=9$) with data from path 186 and rows 16 and 17 (acquisition date: June 10, 2000). Only cloud-free pixels are used. The image data were co-registered to the national base maps (with soil strata). Tomppo and Halme (2004) give more details. Note, however, that we used all nine bands as opposed to eight (1,…,5,7,8,9), as normally practiced in Finland. Our de-correlation of the ancillary variables justifies inclusion of all nine bands.

Regional data from the 9th Finnish national forest inventory of Finland and land representative (surrounding) of the forest in the AOI were used as the reference set. A total of 4252 reference pixels, each with an inventory plot providing the desired attributes (DBH, BA, and VOL), came from the mineral soil stratum (MIN) and 1219 reference pixels came from the peatland stratum (PEAT). All were inventoried during the summer of 2000 with averages (±1 std. dev.) on MIN soils of 15.9 (±10.7) m^2 ha^{-1} (BA), 113.9 (±100.9) m^3 ha^{-1} (VOL) and 14.3 (±10.9) cm (DBH). Peatland stratum averages (±1 std. dev.) were 13.3 (±8.7) m^2 ha^{-1} (BA), 79.8 (±73.3) m^3 ha^{-1} (VOL), and 11.8 (±8.6) cm (DBH). No reference plot straddled a stand boundary. A separate analysis was carried out with plots that were allowed to straddle a stand boundary. In this second reference set there were 1492 units on peatland and 5305 on mineral soils. Averages for this second reference set were 1%–5% higher on MIN and 4%–7% higher on PEAT. The ancillary data were as detailed for the target set. A systematic cluster sampling design is applied in the inventory. One cluster in North Karelia contains either 18 (temporary clusters) or 14 (permanent clusters) field plots that are located along a rectangular tract, 300 m apart. The field plot cluster was a half rectangle in South Savo with 14 and 10 plots respectively and with a plot distance of 250 m. Field plots were geo-located with a GPS system. Trees were selected using PPS-sampling, the inclusion probability of a tree being proportional to its cross section area at a height of 1.3 m. A basal area factor of 2 was applied. No reference unit was inside the AOI. Field data were co-located to image pixels by a multi-criterion optimization approach within a 7×7 pixel window. Details of this procedure are in Halme and Tomppo (2001).

An intensive forest inventory of the AOI conducted in the summer of 2000 with a total of 453 plots provided BA and VOL benchmark data for the k_{nn} predictions. Probability-based estimates of the relative error of BA, and VOL for the AOI are small (3.0%–3.4%, Katila and Tomppo (2006)).

4. Results

4.1. FI1 and FI2

The k_{nn} predictions of mean BA, VOL, and TPH over pixels in the target set tracked fairly closely the averages obtained from the probability-based samples. For $k=7$ the deviations as a percent of the inventory averages varied from 1.8% to 3.1%, depending on attribute and data set (Tables 1 and 2). These deviations remained almost constant for higher values of k, but increased slightly for lower values of k. The k_{nn} predictions for all attributes, with the exception of TPH in FI2, were within estimated 95% probability-based confidence intervals for the corresponding inventory estimates. We should, therefore, accept the prospect of a slight positive bias in k_{nn} predictions of TPH in FI2. $\widehat{\mathrm{MSE}}_{CV}(\tilde{y}_i)$ declined as expected by adding an additional nearest neighbour, but the rate of decline became less than 0.01 for $k\geq 7$. We, therefore, chose $k=7$ for reporting results. $\widehat{\mathrm{RRMSE}}\left(\tilde{T}_y\right)$ was, for $k=7$, in the range from 2.5% to 3.7% for the FI1 data and 2.1% to 2.6% for the FI2 data. For FI1 it was, in relative terms, about 20% higher than the naive $\widehat{\mathrm{RRMSE}}_{CV}(\tilde{y}_i)\times N^{-0.5}$ that would be obtained by a prorating of the crossvalidation MSE to a sum of N k_{nn} predictions in disregard of any covariance between predictions (Table 1). For FI2 the corresponding difference was about 10%

Table 2
FI2 inventory and k_{nn} predictions

	Reference set			Target set		
	BA $m^2 \times ha^{-1}$	VOL $m^3 \times ha^{-1}$	TPH ha^{-1}	BA $m^2 \times ha^{-1}$	VOL $m^3 \times ha^{-1}$	TPH ha^{-1}
Probability-based average $\left(\hat{\bar{y}}\right)$	15.2	85.2	422	15.5	85.6	426
k_{nn} prediction of average $(\tilde{\mu}_y)$	15.1 (−0.5)	84.6 (−0.7)	421 (−0.3)	15.9 (3.0)	88.9 (3.8)	442 (3.8)
$\widehat{\mathrm{RRMSE}}\left(\hat{\bar{y}}\right)\%$[a]	1.7	2.0	1.6	2.1	2.3	1.9
$\widehat{\mathrm{RRMSE}}_{CV}(\tilde{y}_i)\times N^{-0.5}\%$	1.5	1.8	1.4	2.0	2.3	1.8
$\widehat{\mathrm{RRMSE}}(\tilde{y}_i)\times N^{-0.5}\%(cf.4)$	1.5	1.8	1.4	1.3	1.6	1.5
$\widehat{\mathrm{RRMSE}}\left(\tilde{T}_y\right)\%(cf.8)$				2.2	2.6	2.1

Number of pixels (subplots) is 3117 (60%) in the reference set and 2078 (40%) in the target set. All k_{nn} results are for $k=7$. Numbers in parentheses are the relative deviation between a k_{nn} prediction, and the estimate from a probability sample: $\left(\tilde{\mu}_y-\hat{\bar{y}}\right)\times\hat{\bar{y}}^{-1}\times 100$.
[a] Estimated as per Cochran (1977 p 250) for a single stage cluster sampling with unequal number of units (subplots) per sample-location (cluster).

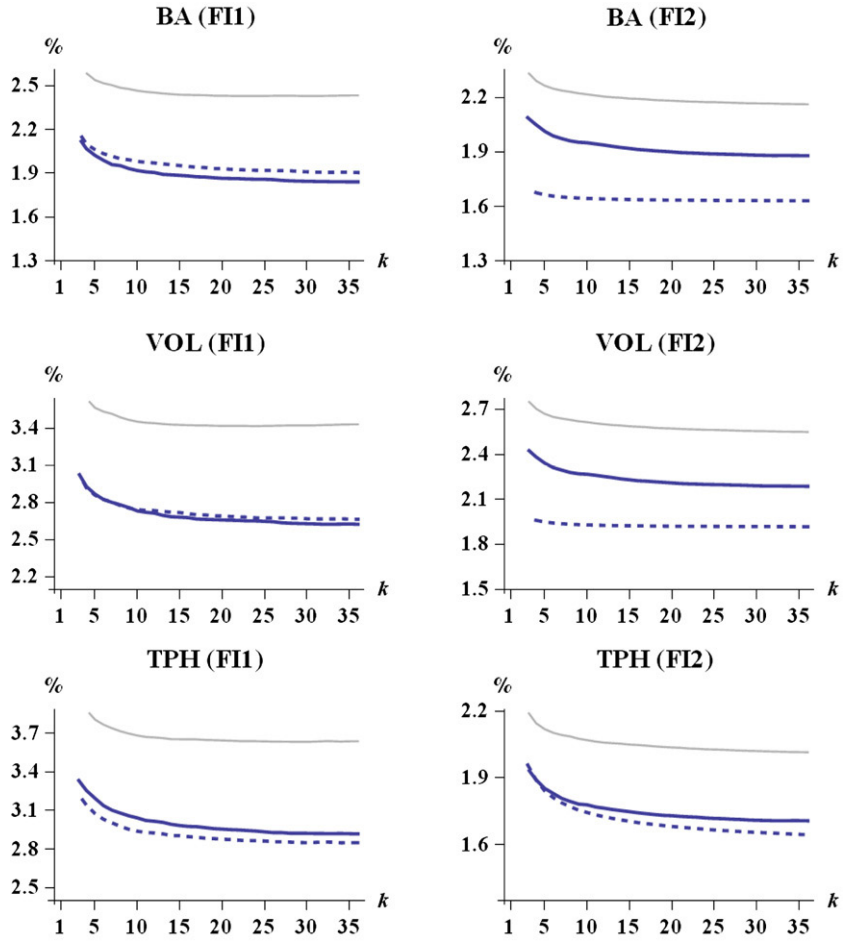

Fig. 1. FI1 and FI2 trends in RRMSE of k_{nn} target set predictions plotted against k. Gray line: $\widehat{\mathrm{RRMSE}}\left(\tilde{T}_y\right)$. Dashed line: $\widehat{\mathrm{RRMSE}}\left(\tilde{y}_i\right)\times N^{-0.5}\%$. Full black line: $\widehat{\mathrm{RRMSE}}_{CV}\left(\tilde{y}_i\right)\times N^{-0.5}\%$.

(Table 2). Since a crossvalidation MSE estimate ignores the covariance between k_{nn} predictions, it is necessarily downward biased if prorated to a prediction for an AOI.

A comparison of $\widehat{\mathrm{RRMSE}}(\tilde{y}_i)$ obtained from Eq. (4) to their empirical counterparts $\widehat{\mathrm{RRMSE}}_{CV}(\tilde{y}_i)$ shows, for FI1, that the former is about 25% below the latter. This contrasts with a much smaller deviation of just 3% for the reference set (Table 1). In FI2 the underestimation was in the neighbourhood of 30% to 50% for the target set but less than 3% for the reference set (Table 2). We surmise that a main portion of the apparent underestimation arises from set differences in conditional expectations ($\mu_{y|\mathbf{x}}$) and variances (Γ_i). Since our variance estimators predict the expected variance of a unit, the predicted variance will, on average, be downward biased by having ignored the variances of the squared deviations (see Appendix 2). Trends in $\widehat{\mathrm{RRMSE}}(\tilde{y}_i)$ $\widehat{\mathrm{RRMSE}}_{CV}(\tilde{y}_i)$, and $\widehat{\mathrm{RRMSE}}\left(\tilde{T}_y\right)$ across k are shown in Fig. 1 for target sets FI1 and FI2. The marginally decreasing effect of k is clear and the practical significance of increasing k beyond 7 seems negligible.

If each reference set is viewed as a representative probability sample from an AOI, then a probability (viz. design-based) estimate of RRMSE would be in the range of 1.6% to 3.2% (Tables 1 and 2) which fairly closely matches corresponding k_{nn} crossvalidation estimates.

4.2. FI3

Estimates of pixel averages and of RRMSE of BA, VOL and DBH in the AOI are given in Table 3 for $k=10$ by strata (PEAT, MIN) and combined. Increasing k beyond 10 had only a small positive effect on $\widehat{\mathrm{MSE}}_{CV}(\tilde{y}_i)$ as can be taken from Fig. 2. The strata and combined strata k_{nn} predictions compare very well with the estimates obtained from the independent intensive probability-based forest inventory. Differences are in the order of 4% on PEAT and below 0.5% on MIN. The magnitudes of these differences are well within a 95% confidence interval around the inventory results; hence we can, with a reasonable

Table 3
FI3 k_{nn} predictions of AOI averages of BA, VOL, and DBH

		Ba m^2ha^{-1}	VOL m^3ha^{-1}	DBH cm
PEAT ($N=2976$)				
k_{nn}	$\tilde{\mu}_y$	15.3(15.5)	93.9(97.2)	13.8(14.3)
	$\widehat{\mathrm{RRMSE}}(\tilde{y}_i)\times N^{-0.5}\%$	2.0	3.0	2.3
	$\widehat{\mathrm{RRMSE}}\left(\tilde{T}_y\right)\%$	9.0	16.4	11.9
INV	$\hat{\bar{y}}$	16.0	98.6	
	$\mathrm{RRMSE}\left(\hat{\bar{y}}\right)\%$	4.8	6.0	
MIN ($N=14{,}644$)				
k_{nn}	$\tilde{\mu}_y$ $\left(\mathrm{RRMSE}\left(\hat{\bar{y}}\right)\%\right)$	18.1(17.9)	133.0(133.9)	16.3(16.7)
	$\widehat{\mathrm{RRMSE}}(\tilde{y}_i)\times N^{-0.5}\%$	0.5	0.6	0.5
	$\widehat{\mathrm{RRMSE}}\left(\tilde{T}_y\right)\%$	7.9	12.5	5.4
INV	$\hat{\bar{y}}$	18.2	132.8	
	$\mathrm{RRMSE}\left(\hat{\bar{y}}\right)\%$	2.7	2.8	
PEAT + MIN ($N=17{,}620$)				
k_{nn}	$\tilde{\mu}_y$	17.6(17.5)	126.4(127.7)	15.9(16.3)
	$\widehat{\mathrm{RRMSE}}(\tilde{y}_i)\times N^{-0.5}\%$	0.4	0.6	0.5
	$\widehat{\mathrm{RRMSE}}\left(\tilde{T}_y\right)\%$	7.3	11.5	5.0
INV	$\hat{\bar{y}}$	17.8	127.0	
	$\mathrm{RRMSE}\left(\hat{\bar{y}}\right)\%$	2.4	2.6	

Results are for $k=10$. Numbers in parentheses are k_{nn} predictions obtained with no exclusion of stand-overlapping reference plots. Estimates derived from an intensive probability-based forest inventory (INV) with 448 plots (349 on MIN and 89 on PEAT) are for comparison.

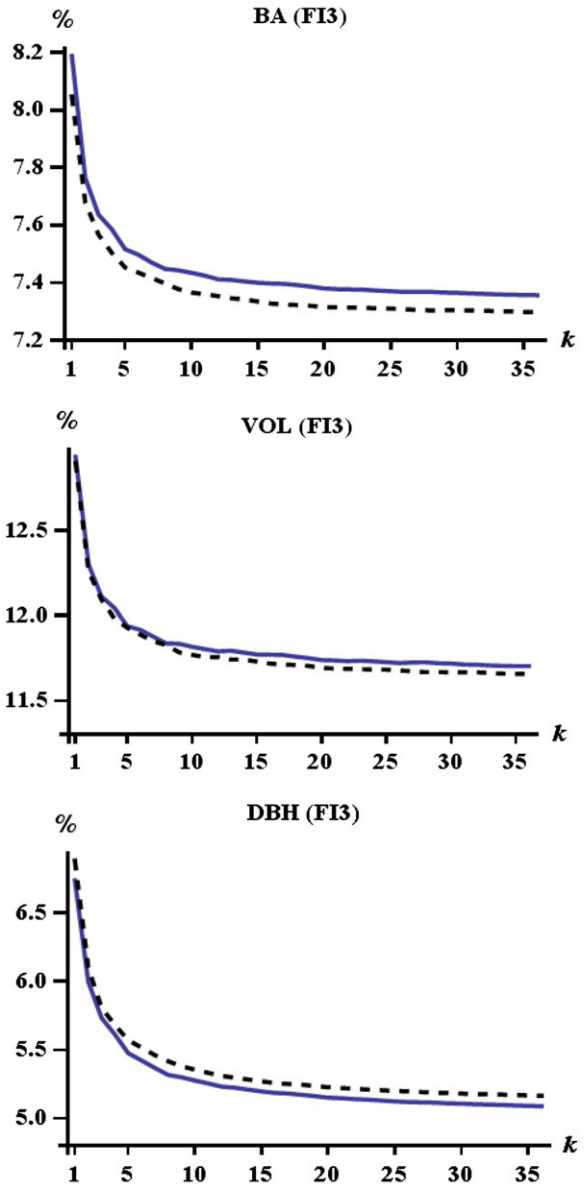

Fig. 2. FI3 trends in $\widehat{\mathrm{RRMSE}}\left(\tilde{T}_y\right)$. Full line: with exclusion of strand-straddling subplots. Dashed line: no exclusion of stand-straddling subplots.

level of confidence, assume that our k_{nn} predictions are approximately unbiased. Relative root mean square error estimates varied from a low of 5% for DBH in the combined stratum (PEAT+MIN) to a high of 16% for VOL on PEAT. The RRMSE in the MIN stratum are lower than those for the PEAT stratum (Table 3). Relative root mean square error for a k_{nn} prediction of an AOI average is 4 and 20 times greater than the naive estimate $\widehat{\mathrm{RRMSE}}_{\mathrm{CV}}(\tilde{y}_i)\times N^{-0.5}$ (Table 3). A fairly strong (average) spatial autocorrelation of k_{nn} predictions in the AOI accounts for a large portion of this difference, as do the covariances of predictions due to sharing of reference predictors.

Allowing reference plots to straddle stand boundaries had little effect on RRMSE (Fig. 2), but some of the k_{nn} predictions of unit averages were sensitive to this screening of the reference plots. The most pronounced difference was for VOL in the PEAT stratum. Without a screening the k_{nn} predictions would be almost 4% higher, in agreement with the fact that the MIN stratum is the dominant stratum in the region and the volumes are generally higher in this stratum.

Relative root mean square error estimates for a single peatland k_{nn} prediction were close to their empirical counterparts obtained from the crossvalidation procedure (Fig. 3). However, our k_{nn} predictions appear to be too high on mineral soils, especially with regards to VOL, as the empirical counterparts are 30% to 50% lower. Since the estimation and prediction procedures were the same for both strata it is not immediately clear what is behind this strata-specific behavior.

5. Discussion

Ancillary remotely-sensed data have been exploited successfully by probability-based forest inventories in efforts to improve precision of regional or national estimates of forest resource attributes (Corona et al., 2002; Holmstrom et al., 2002; Lefsky et al., 2005; McRoberts et al., 2005; Meng et al., 2007; Opsomer et al., 2007). Imagery with a spatial resolution between 5 and 30 m are popular choices for the ancillary variables since their resolution is comparable to that of the plot sizes. Among the many methods applied in a multi-source integration of data, the *k*-nearest neighbour technique has become popular. It has gained interest particularly among forest inventory groups because it is more suitable for forest inventory purposes than a classification approach. With k_{nn}, it is possible to compute new area representatives or plot expansion factors for the field plots, making it possible to predict all interesting forest parameters simultaneously. It is possible also to utilise field plot data outside the area of interest akin to a synthetic estimation method. The k_{nn} is also relatively easy to implement, thus increasing its popularity outside of traditional forest inventory applications. It can be used for pixel-level and areal predictions of forest attributes, even when a large number of forest attributes (>20) have to be predicted for a large number of pixels (Finley et al., 2006; Katila, 2006; LeMay et al., 2008; Mäkela & Pekkarinen, 2004).

Until recently, the absence of an estimator of the MSE of k_{nn} predictions applicable to an aggregate of pixels (AOI) has been a recognized problem. The naive MSE, obtained in a crossvalidation procedure applied to a reference set (Maselli & Chiesi, 2006 for example), can only serve as a first approximation to the pixel-level uncertainty. One cannot simply prorate an MSE estimate for a single pixel to an MSE for their sum as if predictions were independent. Most importantly, the same reference unit is usually chosen repeatedly as a k_{nn} predictor for several target units, which generates a non-trivial positive covariance among predictions, especially when the number of reference units is modest or *k* is large. As well, in environmental surveys a spatial autocorrelation would add to the covariance among predictions (McRoberts et al., 2007). Consequently, the MSE from a crossvalidation, if used to convey the error of a k_{nn} prediction for an AOI, could be seriously misleading (too small). Koistinen et al. (2008) provides a convincing example of the difference between pixel-level and region-level estimates of MSE. Our study confirmed that the MSE of a k_{nn} prediction for an AOI can be several times larger than an MSE of a pixel-level prediction prorated to the AOI level under the assumption of independent pixels. Results for the spatially compact AOI in FI3 are, in this regard, more pertinent than those for the pseudo AOIs in FI1 and FI2. In FI1 and FI2 the AOIs do not form a spatially contiguous set of pixels and the spatial covariance is effectively limited to pixels from a single FIA sample location.

Kim and Tomppo (2006) proposed a variogram-based MSE estimator for pixel-level k_{nn} predictions that – in principle at least – could be extended to aggregates of pixels. However, a generalization of the variogram model would mean replacing the average pixel-level error by a pixel dependent error. McRoberts et al. (2007) were, to our knowledge, the first to propose a model-based k_{nn} estimator of the variance of predictions for single pixels as well as for predictions pertaining to means involving multiple pixels in an AOI. Our super-population model is identical to theirs, but we chose to include models for μ and Γ instead of relying on assumptions that may be realistic only in some cases but not in others. A corroboration of assumptions is always incumbent on users of a model-based estimator of variance. We acknowledge, however, that the statistical properties of a k_{nn} prediction are so

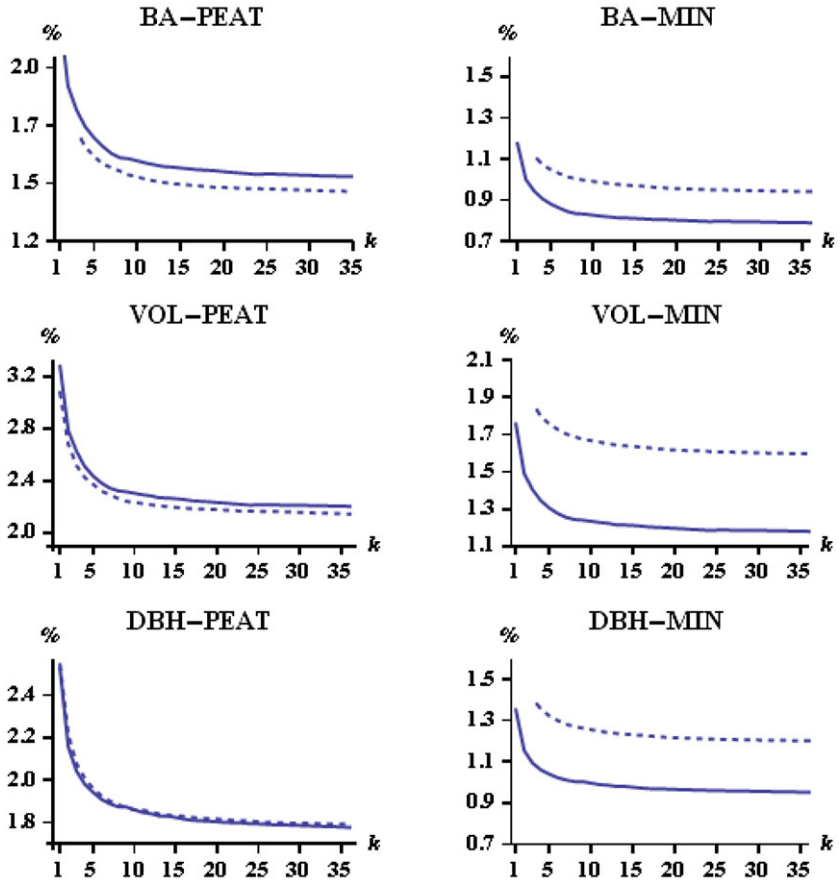

Fig. 3. FI3 trends in RRMSE of k_{nn} reference set predictions plotted against k. Dashed line: $\widehat{RRMSE}(\bar{y}_i)\times N^{-0.5}\%$. Full black line: $\widehat{\mathrm{RRMSE}}_{CV}(\bar{y}_i)\times N^{-0.5}\%$.

complex that only a partial corroboration is feasible. For example, a comparison of our estimates of variance Γ_i to those proposed by McRoberts et al. (2007) showed a good agreement for large k (>25) but a poor agreement for small values of k (<7). The agreement was favorable when the true attribute value was located inside the convex hull spanned by the attribute values of the k-nearest neighbours. For k<7 the chance that the true attribute value would be in the convex hull, was less than 0.75 in all cases (FI1–FI3) and for all three attributes. For k>25 the chance was over 0.90 and the agreement was stronger.

We mentioned that an important attraction of k_{nn} is the elimination of the travails of finding a suitable model. Compared to the McRoberts et al. (2007) approach, our estimator needs four additional models. As the number of models and parameters to be estimated (predicted) from the reference set increases, the risk of overfitting and producing overly optimistic estimates of MSE becomes an important issue and puts a lower limit on the size of the reference set. For the modelling approach taken here, we would cautiously set this limit at 300 reference units. We already saw results that suggest overfitting even with a much larger reference set. The case of model-based underestimation of uncertainty could also be traced to the fact that model-uncertainty is not taken into account. A different reference set would definitely produce a variation in estimates of model parameters. Hastie and Tibshirani (1990) have developed an MSE estimator for a non-parametric smooth (our g-functions) that appears applicable to our SIM.

If it turns out that a model-based prediction of the expected value of y_i is needed for a model-based k_{nn} MSE estimator, a user might be tempted to dispense with the k_{nn} step – at least in the univariate case – and rely on model predictions instead. Model predictions with a contextual post-estimation addition of stochastic errors would then restore the expected range of variability and correlation structures in the predictions. A non-parametric model approach is intuitively appealing (Koistinen et al., 2008), but the performance generally deteriorates fast as the dimension of the ancillary variables increases relative to the size of the reference set (Scott, 1992). The SIM approach was developed to address this 'curse of dimensionality' problem and conventional diagnostic tools can be used to assess goodness-of-fit. The SIM also eliminates a potentially difficult search for a suitable model and facilitates the estimation of an attribute variance function. When the relationship between the attribute and the ancillary variables is locally linear the SIM model should be near optimal (Carroll et al., 2007; Hristache et al., 2001). We tried to find multivariate linear regression models with a smaller residual sum of squares, but were unsuccessful. Multidimensional scaling may offer similar advantages as SIM (Young & Hamer, 1994).

Adherents of the k_{nn} technique may rightfully question the need for an extensive modelling effort. We must learn from experience and studies what is needed and what we can dispense with. To this end we recomputed our results for FI2 and FI3 with the assumptions adopted by McRoberts et al. (2007). Our MSE estimates at the pixel-level were consistently slightly greater than those of McRoberts et al. (2007). In cases of FI2 and FI3-PEAT our estimates are 0.7%–1.9% above corresponding McRoberts et al. (2007) estimates, however for FI3-MIN the differences were in the range of 0.1% to 0.2%. We believe that a large portion of the discrepancy stems from an underestimation of the variance of individual attribute values in the McRoberts et al. (2007) approach. At the AOI level the agreement of the results was generally good in the case of FI2 (differences were less than 0.3%), but results from FI3 point towards systematic difference of a different kind. Our AOI error estimates for MIN were 1.3%–6% less than corresponding McRoberts et al. (2007) estimates. Error estimates for PEAT were mixed with no clear trend, as differences between McRoberts et al. (2007) and ours were from −2.4% (VOL) to +0.1% (BA). We explain the MIN differences by our inclusion of the correlation among the k_{nn} reference units for the

computation of the covariance of k_{nn} predictions. Had we used only the spatial correlation, our results would have been more similar. Our estimates of bias-squared made only a negligible contribution to the MSE.

Validating a k_{nn} MSE estimate of $\tilde{T}_y$ for an AOI level is difficult. The example of FI3 highlights the value of having test areas for which the totals of an attribute of interest have been estimated with a high degree of accuracy. The analyst can assess if bias is a problem and also gauge whether a model-based MSE estimate is realistic or not. It is the choice of k, the weighting scheme, and the strength of the association between the attribute and the ancillary variables that ultimately governs the MSE. In data-rich applications a region-based cross-validation procedure (Hall, 2005) as tried by Koistinen et al. (2008) is a promising option for benchmarking areal MSE estimates.

McRoberts et al.'s (2007) estimator of variance and our proposed model-based approach towards a k_{nn} MSE estimator are but two attempts to present estimators with attractive properties of consistency and robustness. We fully recognize that further improvements are needed and we strongly encourage their developments.

6. Conclusions and recommendations

Formulating a credible and robust model-based k_{nn} estimator of MSE is a complex task due to the non-parametric (distribution free) nature of the k_{nn} procedure, and a reliable estimation requires a relatively large number of reference units. In k_{nn} applications with a large set of reference units representative of the target set, bias seems to be a minor issue. The MSE of an areal prediction is dominated by the correlation of k_{nn} predictions due to the repeated use of the same reference units in different predictions, and possibly a spatial correlation of reference attributes values. An MSE derived from a leave-one-out-crossvalidation will seriously underestimate the precision of areal estimates. When individual predictions are based on a large and representative reference set the risk of bias should be low and give way to simpler estimators of variance (McRoberts et al., 2007) after due corroboration of key assumptions.

Appendix A

A.1. Estimating the unit level expected value μ

The expected value μ_i of y_i was estimated – via a non-parametric regression (g) using the information available in the q ancillary variables – as $\hat{\mu}_i=\hat{g}(x_i^*)$ where x_i^* is a scalar obtained by a SIM transformation of $\mathbf{x}_i$ (Härdle et al., 1993).

Unless the reference set is very large a g estimated with even a moderate number of explanatory variables will suffer from the curse of dimensionality. In clear text: the data supporting the estimate of g will be sparse through most of the prediction domain. We, therefore, opted for a SIM in order to convert, for each attribute of interest, the q ancillary variables into a single index (scalar) suitable as a predictor in a non-parametric regression (g). Conversion of $\mathbf{x}_i$ to x_i^* via SIM and the estimation of g are linked in a single iterative procedure that begins by a decorrelation of the q ancillary variables followed by a scaling to the interval [0,1]. Let $\mathbf{z}_i$ denote the decorrelated and scaled transform of $\mathbf{x}_i$. It is assumed that the relationship between y_i and $\mathbf{z}_i$ is locally linear but globally nonlinear. The decorrelation was done via a Cholesky transform (Rencher, 1995 p 29). For any function $g(x^*)$ we seek a length q row vector $\boldsymbol{\theta}$ so that the gradient of g at every point x_i^* is proportional to $\boldsymbol{\theta}$ in accordance with the assumed local linear relationship

$$\nabla g\left(x_i^*\right)=\partial g\left(x_i^*\right)/\partial x_i^*=\boldsymbol{\theta}g'(\boldsymbol{\theta}.\mathbf{z}_i), \tag{A1}$$

thus, the average gradient in A2 is a linear function of the unknown g.

$$\boldsymbol{\beta}\equiv n^{-1}\boldsymbol{\theta}\sum_{i=1}^{n} g'(\boldsymbol{\theta}.\mathbf{x}_i) \tag{A2}$$

If g is smooth and approximately linear within a small neighbourhood around $\mathbf{z}_i$ then we can obtain estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ from

$$\hat{\boldsymbol{\beta}}=n^{-1}\sum_{i=1}^{n}\widehat{\nabla g}\left(x_i^*\right) \text{ and } \hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\beta}}\times|\hat{\boldsymbol{\beta}}|^{-1} \tag{A3}$$

where $\widehat{\nabla g}(x_i^*)$ is an initial estimator of the gradient and $|\cdot|$ denotes the length (norm) of a vector. An iterated ($m=1,2,\ldots$) local least squares algorithm was used to find simultaneously estimates of $g(x_i^*)$ and $\boldsymbol{\theta}$

$$\begin{pmatrix}\hat{g}_m\left(x_j^*\right)\\ \widehat{\nabla g}_m\left(x_j^*\right)\end{pmatrix}=\left\{\sum_{j'=1}^{n}\begin{pmatrix}1\\ \mathbf{z}_{jj'}\end{pmatrix}\begin{pmatrix}1\\ \mathbf{z}_{jj'}\end{pmatrix}'K\left(\frac{|\mathbf{S}_m\mathbf{z}_{jj'}|^2}{h_m^2}\right)\right\}^{-1}\times\sum_{j'=1}^{n} y_j\begin{pmatrix}1\\ \mathbf{z}_{jj'}\end{pmatrix}K\left(\frac{|\mathbf{S}_m\mathbf{z}_{jj'}|^2}{h_m^2}\right),\ m=0,1,2,\ldots \tag{A4}$$

where $\mathbf{z}_{jj'}=\mathbf{z}_j-\mathbf{z}_{j'}$, K is the bisquare kernel (Hallin et al., 2004), h is a kernel bandwidth, ($h_m=e^{1/(3\min(4,q)}\times h_{m-1}$), $\mathbf{S}_m=(\mathbf{I}_{q\times q}+\rho_m^{-2}\hat{\boldsymbol{\beta}}_{m-1}\hat{\boldsymbol{\beta}}'_{m-1})^{1/2}$ is a $q\times q$ matrix that shrinks $\mathbf{z}_{ij}$ in directions of $\hat{\boldsymbol{\beta}}_m$ and stretches it in directions orthogonal to $\hat{\boldsymbol{\beta}}_m$, and $\rho_m=e^{-1/6}\rho_{m-1}$ is an adaptive tuning constant controlling the magnitude of the shrinkage of $\mathbf{z}_{ij}$. After each iteration $\hat{\boldsymbol{\beta}}_m$ and $\hat{\boldsymbol{\theta}}_m$ were obtained as in (A3) from the current estimate of the average gradient. Starting values were $h_0=(\log(n)/n)^{-\min(0.25,q^{-1})}$, $\rho_0=1$, and $\hat{\boldsymbol{\beta}}_0=\hat{\boldsymbol{\beta}}_{OLS}$. Iterations were stopped when the condition $\rho_k\le n^{-1/3}$ was first satisfied, usually after $2\times\log(n)$ iterations. A goodness-of-fit test for the estimated function g was done according to Xia et al. (2004).

Scatter-plots in Figs. A1 and A2 of y against x^* show the fitted non-parametric functions $\hat{g}$. Estimates of $\boldsymbol{\theta}$ are available from the corresponding author upon request.

A.2. Estimating the unit level variance Γ of y_i

The variance of y_i is $\Gamma_i=E(y_i-\mu_i)^2=Var(\delta_i)=E(\delta_i^2)$. Thus, to estimate Γ_i, we need an estimate of δ_i which, in turn, was obtained from the reference data as $\hat{\delta}_j=y_j-\hat{g}(x_j^*)=y_j-\hat{\mu}_j$, $j=1,\ldots,n$. The reference data suggested the following nonlinear relationship between x_j^* and Γ_j: $\Gamma_j=e^{a+bx_j^*}$. Estimates of Γ_j were obtained by nonlinear least squares methods with $\hat{\delta}_j^2$ as the dependent variable (Davidian & Carroll, 1987). Estimates of six variance functions are in Figs. A1 and A2. Our estimates of Γ_j do not fully account for the expected variance of δ_j^2 nor the leverage of δ_j (Section 3.1). As a consequence we recognize that our estimates may be biased (downward).

In the multivariate case with p attributes we get $\boldsymbol{\Gamma}_i=(\mathbf{y}_i-\boldsymbol{\mu}_i)^T(\mathbf{y}_i-\boldsymbol{\mu}_i)$ a $p\times p$ matrix of variances (diagonal) and covariances in off-diagonal positions (a superscript T denotes the transpose of a matrix). Estimation proceeds by non-linear least squares of a system of simultaneous equations constrained to assure a positive definite covariance matrix.

A.3. Estimating the unit level variance ω of k_{nn} residuals $\tilde{\varepsilon}_i$

The variance of $\tilde{\varepsilon}_i$ depends on the correlation induced by ordering the k reference units by their $\mathbf{X}$-distance to the target unit and on the weights given to these k units. If we assume that the process of ordering of the k-nearest neighbours induces a constant correlation structure on all ordered sets of k reference units (Sarhan & Greenberg, 1962; Stern, 1990), then we have

$$\mathrm{Var}(\tilde{\varepsilon}_i)=\tilde{\omega}_i=\sum_{u(i)=1}^{k}\sum_{v(i)=1}^{k} w_{u(i)}w_{v(i)}\times r_{u,v}\,\Gamma_{u(i)}^{0.5}\Gamma_{v(i)}^{0.5} \tag{A5}$$

where $r_{u,v}$ is the expected correlation among the attribute values of the uth and vth distance-ranked nearest neighbours. We have already detailed estimation of Γ_i in A.2. For an estimation of ω_i we also need an estimate of $r_{u,v}$. Appendix A.4 outlines details of the estimation of $r_{u,v}$.

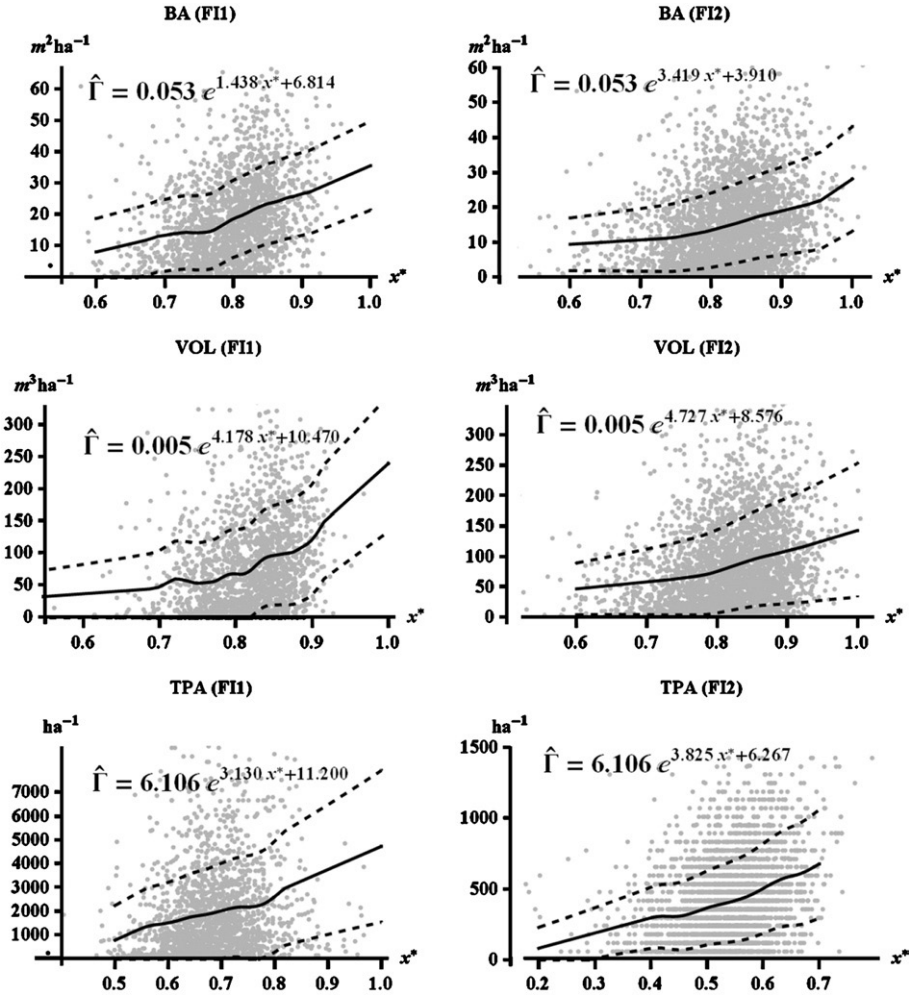

Fig. A1. Scatter plots of y_i (Y={BA, VOL, TPH}) versus x^* in reference sets of FI1 and FI2. Estimates of the conditional expectation of y_i ($\hat{\mu}_i|x^* = \hat{g}(x^*)$ are indicated by a full line. Estimates of $\mu_i|\, x^*$ plus/minus one estimated standard deviation of the mean are indicated by dashed lines.

In the multivariate (p) case the residual covariance matrix becomes

$$\mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_i, \tilde{\boldsymbol{\varepsilon}}_i) = \boldsymbol{\Omega}_i = \sum_{u(i)=1}^{k} \sum_{v(i)=1}^{k} w_{u(i)} w_{v(i)} \times \boldsymbol{\Gamma}_{u(i)}^{T0.5} \mathbf{R}_{u,v} \boldsymbol{\Gamma}_{v(i)}^{0.5} \quad \text{(A6)}$$

Where $\mathbf{R}_{u,v}$ is the $p \times p$ covariance matrix of correlation coefficients among attribute values of the uth and the vth nearest neighbour.

A.4. Correlation r of k-nearest neighbours

In a k_{nn} prediction the reference units are ordered by their distance in **X**-space to the target unit, with the k-nearest units selected for prediction purposes. This ordering process induces a correlation among the y-values of the selected reference units (Harter, 1970). The correlation will depend on both the distribution of y and the ordering process. Since y-values are ordered with respect to a distance measure applied to ancillary variables, the correlation structure induced by this process is analytically intractable. One should expect the correlation structure to also depend on the actual **X**-distances associated with a specific set of k reference units and possibly also on whether the prediction is for a target unit with an **X**-value outside the support domain of the reference units. In this study, however, we just estimated, for each attribute of interest, the $k \times k$ matrix of Pearson's product moment correlation coefficients $r_{u,v}, u,v = 1,\ldots,k$ from standardized deviations $(\hat{\delta}_{u(j)} \times \Gamma_{v(j)}^{-0.5}, u,v = 1,\ldots,k)$ in a k_{nn} leave-j-out cross-validation procedure applied to the reference set ($j = 1,\ldots,n$). Estimates of $r_{u,v}, u,v = 1,\ldots,k$ for BA and $k = 1,\ldots,7$ are in Tables A1 and A2. The tabled estimates are representative also for other studied variables (VOL, TPH, and DBH).

A.5. Estimating η the correlation between δ_i and $\tilde{\varepsilon}_i$

We obtained an estimate of η from reference set estimates of δ_j and $\tilde{\varepsilon}_j$ in a leave-j-out k_{nn} crossvalidation procedure ($j = 1,\ldots,n$). We estimated η by the Pearson product moment correlation between $\hat{\delta}_j = y_j - \hat{\mu}_j = y_j - \hat{g}(x_j^*)$ and $\hat{\tilde{\varepsilon}}_j = \tilde{y}_j - \sum_{u=1}^{k} \hat{\mu}_{u(j)}$. Estimates of η for $k = 1,\ldots,7$ are in Table A3. In a multivariate setting with p attributes of interest we must estimate a $p \times p$ matrix $\boldsymbol{\Psi}$ of coefficients $\eta(q,q')$, $q,q' = 1,\ldots p$ estimated from $\hat{\delta}_j(q)$ and $\hat{\tilde{\varepsilon}}_j(q')$.

A.6. Estimating $\mathrm{cov}(\tilde{y}_i, \tilde{y}_{i'})$

The expected covariance between two k_{nn} predictions, say $\tilde{y}_i$ and $\tilde{y}_{i' \neq i}$ is determined by the number of reference units they have in common and by the spatial correlation (ρ) between the two sets of reference units. We have

$$\mathrm{cov}\left(\tilde{y}_i, \tilde{y}_{i' \neq i}\right) = \sum_{u=1}^{k} \sum_{u=1}^{k} w_{u(i)} \times w_{v(i')} \times \rho\left(\delta_{u(i)}, \delta_{v(i')}\right) \Gamma_{u(i)}^{0.5} \Gamma_{v(i')}^{0.5} \quad \text{(A7)}$$

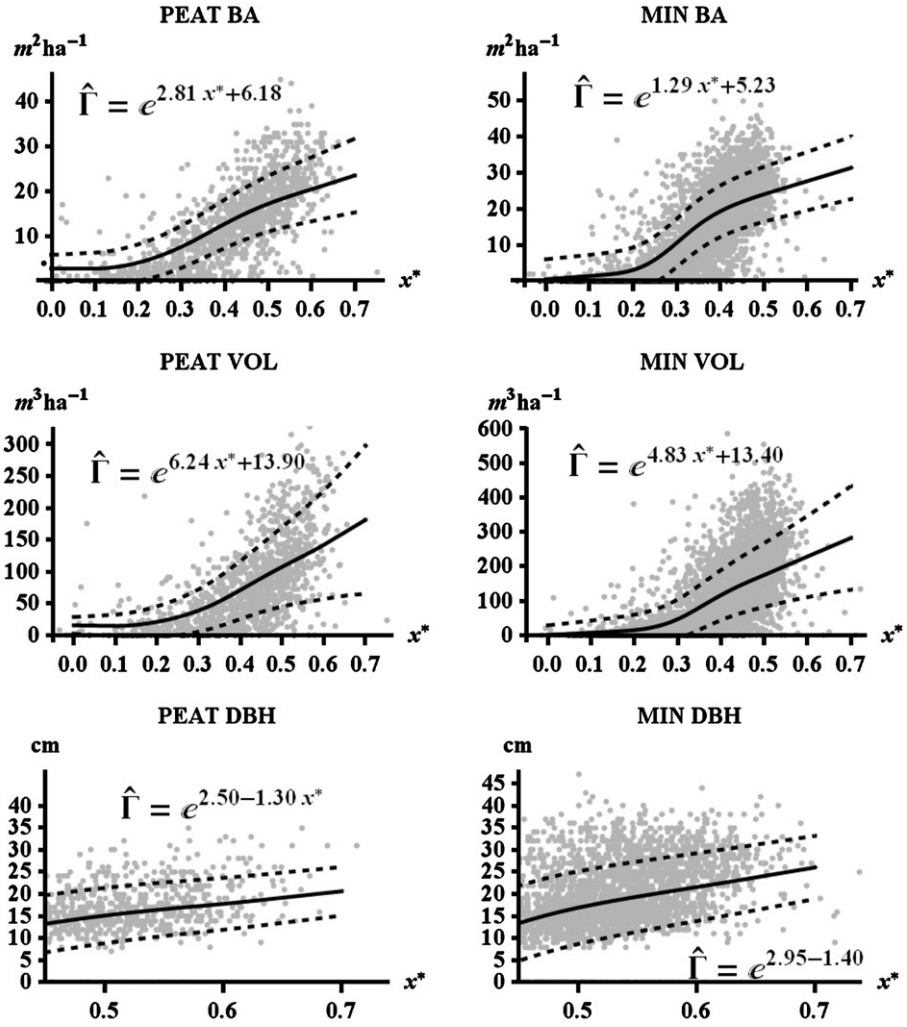

Fig. A2. Scatter plots of y_i (Y={BA, VOL, DBH}) versus x^* in reference sets of FI3. Estimates of the conditional expectations of y_i ($\hat{\mu}_i|x^*=\hat{g}(x^*)$ are indicated by a full line. Estimates of $\mu_i|\,x^*$ plus/minus one estimated standard deviation of the mean are indicated by dashed lines.

where ρ is the expected correlation between deviations (δ) of reference units $u(i)$ and $v(i')$. Details on estimation of ρ are in A.9. For a p-variate attribute we have

$$\mathrm{cov}(\tilde{y}_i, \tilde{y}_{i'\neq i}) = \Xi_{i,i'} = \sum_{u=1}^{k}\sum_{v=1}^{k} w_{u(i)} \times w_{v(i')} \times \Gamma_{u(i)}^{T0.5} \boldsymbol{\theta}_{i,i'} \Gamma_{v(i')}^{0.5} \tag{A8}$$

where $\Xi_{i,i'}$ is a $p\times p$ matrix of distance weighted covariances $cov(\delta(q)_{u(i)}, \delta(q')_{v(i')}), q,q' = 1,\ldots,p$.

A.7. Estimating $cov(y_i, y_{i'})$

From our definition of y_i we have $\mathrm{cov}(y_i,y_{i'}) = \mathrm{cov}(\delta_i,\delta_{i'}) = \Gamma_i^{0.5}\Gamma_{i'}^{0.5}\rho(\delta_i,\delta_{i'})$ where, as above, ρ is the expected correlation between deviations (δ) of reference units $u(i)$ and $v(i')$. See A.9 for details. Extension to the estimation of a covariance between attributes q and q' follows immediately.

A.8. Estimating $cov(y_i, \tilde{y}_{i'})$

The covariance between y_i and $\tilde{y}_{i'}$ depends on whether $i=i'$ or $i\neq i'$. We have

$$\mathrm{cov}(y_i, \tilde{y}_{i'}) = \begin{cases} \eta \times \Gamma_i^{0.5} \times \tilde{\omega}_{i'}^{0.5} \text{ if } i = i' \\ \sum_{u=1}^{k} w_{u(i')} \rho(\delta_i, \delta_{u(i')}) \Gamma_i^{0.5} \Gamma_{u(i')}^{0.5} \text{ if } i \neq i' \end{cases} \tag{A9}$$

where $\eta = corr(\delta_i, \tilde{\varepsilon}_i)$ and ρ a correlation function (see A.9 for details). An extension to the covariance among different attributes is straightforward.

Table A1
Estimated correlation matrix of $100\times r_{u,v}$ for $k=7$ nearest neighbour values of standardized deviations (δ_i) of BA in the FI1 and FI2 reference sets

BA	FI1								FI2							
		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
	$k=1$	100	0	4	−8	2	3	−2	$k=1$	100	0	−3	1	−1	−2	0
	$k=2$	0	100	−2	1	0	−1	−1	$k=2$	0	100	1	−1	1	2	−2
	$k=3$	4	−2	100	−1	−4	1	0	$k=3$	−3	1	100	0	2	5	2
	$k=4$	−8	1	−1	100	−2	−1	−1	$k=4$	1	−1	0	100	4	−4	1
	$k=5$	2	0	−4	−2	100	−3	−1	$k=5$	−1	1	2	4	100	3	2
	$k=6$	3	−1	1	−1	−3	100	5	$k=6$	−2	2	5	−4	3	100	1
	$k=7$	−2	−1	0	−1	−1	5	100	$k=7$	0	−2	2	1	2	1	100

Table A2
Estimated correlation matrix of $100\times r_{u,v}$ for $k=7$ nearest neighbour values of standardized deviations (δ_i) of BA in the FI3 reference set

	PEAT								MIN							
BA		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
	$k=1$	100	5	1	0	3	3	7	$k=1$	100	2	0	1	2	4	3
	$k=2$	5	100	1	−1	5	4	2	$k=2$	2	100	2	4	3	0	1
	$k=3$	1	1	100	6	2	1	4	$k=3$	0	2	100	4	3	0	1
	$k=4$	0	−1	6	100	4	7	4	$k=4$	1	4	0	100	3	2	3
	$k=5$	3	5	2	4	100	−2	3	$k=5$	2	3	1	3	100	3	4
	$k=6$	3	4	1	7	−2	100	1	$k=6$	4	0	2	2	3	100	3
	$k=7$	7	2	4	4	3	1	100	$k=7$	3	1	3	3	4	3	100

A.9. Estimating ρ the correlation of deviations δ_i

Unit-level attribute values can be correlated due to proximity in space, similarity of their ancillary values, or both. It is well known that forest inventory attributes gathered at locations separated by a short distance tend to be more similar than attributes at locations separated by a longer distance (Reed & Burkhart, 1985). Conversely, the assumed association between $\mathbf{x}_i$ and y_i will tend to make two attribute values more similar the closer their associated ancillary values are to each other. Since the ancillary values are fixed (by design) and we use them as predictors of attributes of interest, the relevant correlation for the current estimation problem is the correlation ρ among unit level deviations $\delta_i = y_i - \mu_i$. We shall assume that the (expected) correlation is a function only of the Euclidian distance (h) between the two units involved. Inclusion of additional explanatory variables in the correlation function such as, for examples, elevation and soil type would be straightforward and should be considered where appropriate.

Table A3
Reference set estimates of the correlation coefficient η between deviations δ_i and residuals $\tilde{\varepsilon}_i$

		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
FI1	BA	−0.00	−0.01	−0.00	0.01	0.02	0.02	0.02
	VOL	0.02	0.02	0.02	0.04	0.05	0.05	0.04
	TPH	0.04	0.04	0.02	0.02	0.01	0.01	0.01
FI2	BA	−0.03	−0.02	−0.02	−0.01	−0.01	−0.00	−0.00
	VOL	−0.02	−0.02	−0.02	−0.01	−0.01	−0.01	−0.00
	TPH	−0.02	−0.03	−0.03	−0.02	−0.01	−0.01	−0.01
FI3-PEAT	BA	0.05	0.02	0.03	0.04	0.05	0.06	0.06
	VOL	0.02	0.02	0.03	0.04	0.04	0.06	0.06
	DBH	0.04	0.04	0.04	0.04	0.05	0.05	0.06
FI3-MIN	BA	0.03	0.05	0.05	0.07	0.07	0.06	0.08
	VOL	0.02	0.05	0.05	0.06	0.07	0.06	0.07
	DBH	0.03	0.04	0.04	0.04	0.05	0.05	0.05

Results are shown for $k=1,\ldots,7$ only.

Table A4
Weighted least squares estimates of φ in the model of spatial autocorrelation (ρ) as a function of distance (h): $\rho_t(h)=1-e^{-\varphi/h}$ fitted to reference data sets

Reference set	Attribute(y)	$\hat{\varphi}$	$\widehat{se}(\hat{\varphi})$
FI1	BA	2.20	0.46
	VOL	2.40	0.45
	TPH	5.79	0.38
FI2	BA	3.24	0.53
	VOL	3.05	0.50
	TPH	2.63	0.51
FI3-PEAT	BA	26.98	0.78
	VOL	47.54	0.86
	DBH	31.27	0.65
FI3-MIN	BA	28.71	0.71
	VOL	31.86	0.77
	DBH	6.02	0.73

Approximate standard errors of estimates of φ are obtained by the delta technique (Kendall & Stuart, 1969). Larger φ values indicate a slower decay of autocorrelations with distance.

A model of ρ was estimated from the reference data for each attribute of interest. Earlier studies with similar data (Kim & Tomppo, 2006; McRoberts et al., 2007) suggest a simple exponential decline of the correlation as distance (h) increases. Attribute (t) specific nonlinear weighted least squares estimates of the model in I1 were obtained for each data set and p attributes of interest

$$\rho_t(h[i,i']\,|\,h[i,i']>0) = 1-e^{-\varphi_t/h[i,i']}\,,\ \rho(0)\equiv 1\,,\ t=1,\ldots,p \qquad \text{(A10)}$$

where $h[i,i']$ is the Euclidian spatial distance between units i and i'. Weights were inversely proportional to distance in order to counter any potential linear spatial trends in y, and because the precision of predictions over shorter distances is more important than predictions over longer distances (Stein, 1999). Correlation models were estimated from a random sample of about 300,000 distinct pairs (j,j') of reference units using the standardized cross-product $\hat{\delta}_j\times\hat{\delta}_{j'}\times\hat{\Gamma}_j^{-0.5}\hat{\Gamma}_{j'}^{-0.5}$ as the dependent variable (Cressie, 1989 p 67).

Estimating the correlation between deviations $\delta_i(q)$ and $\delta_{i'}(q')$, $q, p' = 1,\ldots,p$ in the multivariate case follows the principles of the univariate case except for the use of attribute specific deviations.

Least squares estimates of φ are in Table A4. The distance required (on average) to lower the correlation from 1.0 (at distance 0) to 0.05 is approximately $19.5\times\varphi$m or about 40 m–100 m in case of FI1 and FI2, and 120–1000 m in FI3. A more uniform forest structure (species composition, stand structure, management, and stratification by soil type) behind the data from the Finnish site is surmised as main factor (s) behind the marked differences among the US and the Finnish sites.

References

Aha, W. D. (1997). *Lazy learning.* Dordecht: Kluwer.

Alt, H. (2001). The nearest neighbor. *Computational Discrete Mathematics: Advanced Lectures*, *2122*, 13–24.

Bechthold, 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* (pp. 1–85). Southern Research Station, Asheville, NC: USDA Forest Service.

Carroll, R. J., Fan, J., Gijbels, I., & Wand, M. P. (2007). Generalized partially linear single-index models. *Journal of the American Statistical Association*, *92*, 477–489.

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

Cliff, A. D., & Ord, J. K. (1981). *Spatial processes.* London: Pion.

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

Collins, M. J., Dymond, C., & Johnson, E. A. (2004). Mapping subalpine forest types using networks of nearest neighbour classifiers. *International Journal of Remote Sensing*, *25*, 1701–1721.

Corona, P., Chirici, G., & Marchetti, M. (2002). Forest ecosystem inventory and monitoring as a framework for terrestrial natural renewable resource survey programmes. *Plant Biosystems*, *136*, 69–82.

Cressie, N. A. C. (1989). Geostatistics. *The American Statistician*, *43*, 197–202.

David, H. A., & Mishriky, R. S. (1968). Order statistics for discrete populations and for grouped samples. *Journal of the American Statistical Association*, *63*, 1390–1398.

Davidian, M., & Carroll, R. J. (1987). Variance function estimation. *Journal of the American Statistical Association*, *82*, 1079–1091.

D Orazio , M., Di Zio, M., & Scanu, M. (2006). *Statistical matching.* Theory and Practice Chichester: Wiley & Sons.

Finley, A. O., McRoberts, R. E., & Ek, A. R. (2006). Applying an efficient *k*-nearest neighbor search to forest attribute imputation. *Forest Science*, *52*, 130–135.

Flores, L. A., & Martínez, L. I. (2000). Land cover estimation in small areas using ground survey and remote sensing. *Remote Sensing of Environment*, *74*, 240–248.

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*, 251–274.

Glatzer, E., & Muller, W. G. (2004). Residual diagnostics for variogram fitting. *Computers & Geosciences*, *30*, 859–866.

Hall, P. (2005). On confidence intervals for spatial parameters estimated from non-replicated data. *Biometrics*, *44*, 271–277.

Hallin, M., Lu, Z. D., & Tran, L. T. (2004). Local linear spatial regression. *Annals of Statistics*, *32*, 2469–2500.

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

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

Härdle, W., Hall, P., & Ichimura, H. (1993). Optimal smoothing in single-index models. *Annals of Statistics*, *21*, 157–178.

Harter, H. L. (1970). *Order statistics and their use in testing and estimation. Vol II.* Washington D.C.: US Government Printing Office.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models.* London: Chapman & Hall.

Holmstrom, H., Nilsson, M., & Ståhl, G. (2002). Forecasted reference sample plot data in estimations of stem volume using satellite spectral data and the kNN method. *International Journal of Remote Sensing*, *23*, 1757–1774.

Hristache, M., Juditsky, A., & Spokoiny, V. (2001). Direct estimation of the index coefficient in a single-index model. *Annals of Statistics*, *29*, 595–623.

Katila, M. (2006). Empirical errors of small area estimates from the multisource national forest inventory in Eastern Finland. *Silva Fennica*, *40*, 729–742.

Katila, M., & Tomppo, E. (2002). Stratification by ancillary data in multisource forest inventories employing *k*-nearest-neighbour estimation. *Canadian Journal of Forest Research*, *32*, 1548–1561.

Katila, M., & Tomppo, E. (2006). Sampling, simulation on multi-source output forest maps — An application for small areas. In M. E. Caetano, & M. Painho (Eds.), *Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences.* Lisbon: Portugese Geographic Institute.

Kauth, R. J., & Thomas, G. S. (1976). The tasselled 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: Purdue University.

Kendall, M. G., & Stuart, A. (1969). *The advanced theory of statistics.* London: Griffin.

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

Koistinen, P., Holmström, L., & Tomppo, E. (2008). Smoothing methodology for predicting regional averages in multi-source forest inventory. *Remote Sensing of Environment*, *112*, 862–871.

Lefsky, M. A., Turner, D. P., Guzy, M., & Cohen, W. B. (2005). Combining lidar estimates of aboveground biomass and Landsat estimates of stand age for spatially extensive validation of modeled forest productivity. *Remote Sensing of Environment*, *95*, 549–558.

LeMay, V., Meadel, J., & Coops, N. (2008). Estimating stand structural details using variable-space nearest neighbour analyses to link ground data, forest cover maps, and Landsat imagery. *Remote Sensing of Enviroment*, *112*, 2578–2591.

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.

Mäkela, H., & Pekkarinen, A. (2004). Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data. *Forest Ecology and Management*, *196*, 245–255.

Marchant, B. P., & Lark, R. M. (2004). Estimating variogram uncertainty. *Mathematical Geology*, *36*, 867–898.

Maselli, F., & Chiesi, M. (2006). Evaluation of statistical methods to estimate forest volume in a Mediterranean region. *IEEE Transactions on Geoscience and Remote Sensing*, *44*, 2239–2250.

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

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 of forest attributes using the *k*-nearest neighbors technique and satellite imagery. *Remote Sensing of Environment*, *111*, 466–480.

McRoberts, R. E., Wendt, D. G., & Liknes, G. C. (2005). Stratified estimation of forest inventory variables using spatially summarized stratifications. *Silva Fennica*, *39*, 559–571.

Meng, Q. M., Cieszewski, C. J., Madden, M., & Borders, B. E. (2007). K nearest neighbor method for forest inventory using remote sensing data. *Giscience & Remote Sensing*, *44*, 149–165.

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

Opsomer, J. D., Breidt, F. J., Moisen, G. G., & Kauermann, G. (2007). Model-assisted estimation of forest resources with generalized additive models. *Journal of the American Statistical Association*, *102*, 400–409.

Rao, J. N. K. (2003). *Small area estimation.* Hoboken, New Jersey: Wiley & Sons.

Reed, D. D., & Burkhart, H. E. (1985). Spatial autocorrelation of individual tree characteristics in loblolly pine stands. *Forest Science*, *31*, 575–587.

Reese, H., Nilsson, M., Sandström, P., & Olsson, H. (2002). Applications using estimates of forest parameters derived from satellite and forest inventory data. *Computers and Electronics in Agriculture*, *37*, 37–56.

Rencher, A. C. (1995). *Methods of multivariate analysis.* New York: Wiley.

Royall, R. M. (1988). The prediction approach to sampling. In P. R. Krishnaiah & C.R. Rao (Eds.), *Sampling* (pp. 399–413). Amsterdam: Elsevier.

Ruppert, D., Wand, M., & Carroll, R. (2003). *Semiparametric regression.* Cambridge: Cambridge Univ. Press.

Sarhan, A. E., & Greenberg, B. G. (1962). *Contributions to order statistics.* Wiley.

Scott, D. W. (1992). *Multivariate density estimation: Theory, practice and visualization.* New York: Wiley.

Searle, S. R. (1982). *Matrix algebra useful for statistics.* New York: Wiley.

Singh, R., Semwal, D. P., Rai, A., & Chhikara, R. S. (2002). Small area estimation of crop yield using remote sensing satellite data. *International Journal of Remote Sensing*, *23*, 49–56.

Stein, M. L. (1999). *Interpolation of spatial data.* Some theory for kriging. New York: Springer.

Stern, H. (1990). Models for distribution on permutations. *Journal of the American Statistical Association*, *85*, 558–564.

Tomppo, E. (2006). The Finnish multi-source national forest inventory — Small area estimation and map production. In A. Kangas & M. Maltamo (Eds.), *Forest inventory — Methodology and applications* (pp. 195–224). Dordrecht, NL: Springer.

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

Valliant, R., Dorfman, A. H., & Royall, R. M. (2000). *Finite population sampling and inference.* A prediction approach. New York: John Wiley & Sons.

Wu, C. B. (2003). Optimal calibration estimators in survey sampling. *Biometrika*, *90*, 937–951.

Xia, Y., Li, W. K., Tong, H., & Zhang, D. (2004). A goodness-of-fit test for single-index models. *Statistica Sinica*, *14*, 1–39.

Young, F. W., & Hamer, R. M. (1994). *Theory and applications of multidimensional scaling.* Hillsdale, NJ: Erlbaum Associates.

Zhang, L. -C., & Chambers, R. L. (2004). Small area estimates for cross classifications. *Journal Royal Statistical Society, Series B*, *66*, 479–496.

Zhao, Y. L., & Wall, M. M. (2004). Investigating the use of the variogram for lattice data. *Journal of Computational and Graphical Statistics*, *13*, 719–738.