Canadian Journal of Forest Research

OPEN ACCESS | Research Article

Small area estimators in a simulation test

Annika Kangas[a], **Mari Myllymäki**[b], **and Petteri Packalen**[b]

[a]Natural Resources Institute Finland (Luke), Bioeconomy and Environment, Yliopistokatu 6, 80101 Joensuu, Finland; [b]Natural Resources Institute Finland (Luke), Bioeconomy and Environment, Latokartanonkaari 9, 00790 Helsinki, Finland

Corresponding author: **Annika Kangas** (email: annika.kangas@luke.fi)

Abstract

The Finnish National Forest Inventory produces municipality level results either with an indirect model-based K-nearest neighbor (K-NN) estimator or a direct design-based post-stratification estimator. Design-based approach is unbiased, but not always feasible due to low number of field plots. The K-NN estimator is lacking an analytical estimator for the variance. A composite estimator combining the indirect and direct estimates could be an attractive solution. In this article, estimators for small-area estimation are analyzed in a simulation experiment with varying size small areas and varying quality auxiliary data. The potential of estimators is assessed based on the true standard errors and RMSEs in the simulation experiment. Direct estimators and composite estimators work reasonably well with varying quality models, but the performance of indirect estimators is dependent on the quality of the model used. The performance of different estimators also depends on the size of the small areas. Linear models in which the weight of plots outside the target domain is smaller than those within the target domain, performed better than an unweighted model, suggesting that localizing the models for the small areas is beneficial. EBLUP approach also performed well, both in connection of a K-NN model and a linear model.

Key words: K-nearest neighbors, mixed model, area effect, composite estimator, indirect estimator, direct estimator

1. Introduction

Based on the Finnish National Forest Inventory (NFI), forest resources maps are provided using K-nearest neighbor (K-NN) method (Mäkisara et al. 2022), and then utilized in estimating municipality level results. In a similar fashion, also a kernel estimator could be used (Kangas 1996), the main difference between the approaches being that in K-NN the number of neighbors is fixed, while in the basic kernel approach the maximum range from which the neighbors are utilized is fixed. Both approaches result in an indirect (or synthetic) estimator, i.e., an estimator that also uses data from outside of the domain of interest. The main problem with K-NN estimator is that an estimator for mean square error (MSE) is lacking. For the variance of the K-NN estimators, some estimators have been proposed (McRoberts et al. 2007; McRoberts 2012; Breidenbach et al. 2018), but the unbiasedness of an indirect estimator is never guaranteed, and the bias component cannot be estimated either.

The direct design-based estimator (i.e., estimator not utilizing data from outside the domain of interest) would be desirable for NFI at municipality-level, as this approach provides unbiased estimates. An example of such an approach is small-area estimation using post-stratification (PS) (Pulkkinen et al. 2018; Haakana et al. 2020; White et al. 2021; Finley et al. 2024). However, while design-based approach is attractive, it may produce unacceptably large standard errors for municipalities with small sample sizes. In this situation, a composite estimator, a weighted sum of a direct and an indirect estimator, may be an attractive option balancing the pros and cons of both estimators (Ghosh and Rao 1994; Rao and Molina 2015). When no observations at all are available from a municipality, the indirect estimators are the only option.

One possibility for this is to use model-based (MB) inference with either an area-level or a unit-level mixed model including an area-effect. This means assuming constant within-area correlations (Magnussen and Breidenbach 2017; Astrup et al. 2019). A block-kriging estimator with a continuous correlation function is a generalization of that assumption for unit level (Wadoux and Heuvelink 2023). An alternative to the area-effect model is a hierarchical Bayesian model, where the area-effect is replaced with prior information regarding the targeted domains (see White et al. 2021).

Both the kriging and the area-effects model can also be seen as composite estimators, where the initial sample statistic within the small area is adjusted using an EBLUP prediction for the unsampled units (Militina et al. 2007). Note that all composite estimators, like all indirect MB approaches, assume that the small areas have similarities across larger area

(Magnussen and Breidenbach 2017). If we can assume that the similarity is more pronounced in the areas that are close to the domain of interest, we could weigh the observations within the target area or close to it more than observations further away (Lappi 2001) or define a maximum range in a same way as in a kernel estimator. This leads to the question: how large is the area around the domain of interest having similar characteristics and how to determine it in a practical application?

The aim of this paper is to test unit-level small-area estimators based on MB and design-based estimation and their composites with varying sizes of domains and varying quality auxiliary information. Specifically, we aim at demonstrating the trade-offs between the precision and accuracy of the results, which in practical applications is unobservable. This is carried out with a simulation study where the "ground truth" is known.

2. Material

We made a simulation experiment to test the performance of the small-area estimators. For this purpose, we utilized wall-to-wall data on a region of size about 5900 ha, a small part of an earlier airborne laser scanning (ALS) campaign. This test data was available on a grid of 231 824 population elements of size 16 m × 16 m, for which coordinates and 17 ALS features are available. To have a "ground truth", we simulated for each pixel i a volume with $y_i = \exp(\mu_i + e_i)$, where μ_i is the predicted logarithm of volume from an external model and e_i is the simulated random error. The errors were assumed to be autocorrelated and stem from a zero-mean Gaussian random field with exponential semivariogram model having variance $\sigma^2 = 0.0538$, nugget effect $\tau^2 = 0.0292$, and range parameter $\phi = 337$, resulting in a practical range of 1011 m (see details from Kangas et al. 2023).

The external model was based on an independent modelling dataset that had 1044 observations with field-measured values of total plot volume, basal area, mean diameter, and mean height, as well as a set of 190 ALS features (see Tuominen et al. (2017) and Balazs et al. (2022) for details). We modelled the plot-specific $\ln(y)$ using the best seven predictor model estimated with leaps package in R,[1] using the following 17 ALS features: the maximum height of the points, height at which given percentiles (20% last echo, 45%, 65%, 70%, 90%, and 95%) of vegetation points are accumulated (m), proportion of vegetation points relative to all points (%, first and last echo), skewness of the vegetation point heights, proportion of points above mean height, proportion of points having cumulated at 20% of the height from all points (%, last echo), Rumple index, inner volume (Vega et al. 2016), SumEntropy (Haralick et al. 1973) of canopy surface model, and average intensity of ALS echoes.Unless otherwise stated, the features were calculated from the first echoes. The "true" model had residual standard error = 0.232 and multiple $R^2 = 0.897$ (Kangas et al. 2023). This seven-predictor model was used only in the generation of the simulated population, not in the estimators.

[1] Please see Thomas Lumley's R-package "leaps" based on the fortran code by Alan Miller in https://cran.r-project.org/web/packages/leaps/leaps.pdf

3. Methods

3.1. Design-based estimators for small areas

The most basic design-based estimator for the mean in domain j is the Horvitz–Thompson (HT) estimator,

$$(1)\qquad \widehat{\bar{y}}_{\mathrm{HT},\,j} = \frac{1}{A_j}\sum_{i=1}^{n_j}\frac{y_{ji}}{\pi_i}$$

where A_j is the total area of domain j and y_{ji} is the variable of interest for unit i in domain j, n_j is the size of the sample in domain j and π_i is the inclusion probability of unit i. Assuming a simple random sampling (SRS) without replacement from the whole population this inclusion probability is $\frac{n}{N}$, where N is the size of the population with J domains and $N = \sum_{j=1}^{J} N_j$ and $n = \sum_{j=1}^{J} n_j$. The estimator of its variance is

$$(2)\qquad var(\hat{\bar{y}}_{HT},j) = \frac{1}{A_j^2}\sum_{i=1}^{n_j}\sum_{l=1}^{n_j}\frac{\pi_{il}-\pi_i\pi_l}{\pi_{il}}\frac{y_{ji}}{\pi_i}\frac{y_{jl}}{\pi_l}$$

where π_{il} is the joint inclusion probability of cells i and l. When $i = l$, this joint probability is π_i, otherwise it is $\frac{n(n-1)}{N(N-1)}$ (Särndal et al. 1992, pp. 31–32).

In small areas, however, the HT estimators may be very inefficient, and utilization of auxiliary data from remote sensing may be advisable. One very popular approach for small-area estimation is the direct model-assisted (MA) estimation which uses besides the observed data also predictions from a model within each domain j. As a mean estimator, either a difference estimator or a generalized regression estimator can be used. The difference estimator for the mean in domain j is (e.g., Kangas et al. 2016)

$$(3)\qquad \widehat{\bar{y}}_{\mathrm{MA},j} = \frac{1}{A_j}\left(\sum_{i=1}^{N_j}\widehat{y}_{ji} + \sum_{i=1}^{n_j}\frac{e_{ji}}{\pi_i}\right)$$

where $\widehat{y}_{ji}$ is the model prediction in unit i and domain j and $e_{ji} = y_{ji} - \widehat{y}_{ji}$. Its variance estimator (assuming g-weights to be 1 for all i; Särndal et al. 1992, p. 362) is

$$(4)\qquad \mathrm{var}\left(\widehat{\bar{y}}_{\mathrm{MA},j}\right) = \frac{1}{A_j^2}\sum_{i=1}^{n_j}\sum_{l=1}^{n_j}\frac{\pi_{il}-\pi_i\pi_l}{\pi_{il}}\frac{e_{ji}}{\pi_i}\frac{e_{jl}}{\pi_l}$$

3.2. Model-based estimators using a mixed model

For indirect MB estimators we assumed a linear-mixed model

$$(5)\qquad y_{ji} = \mathbf{x}_{ji}\boldsymbol{\beta} + v_j + e_{ji},\ \ j = 1, \ldots, J;\ i = 1, \ldots, n_j$$

where e_{ji} is the random error of unit i in domain j ($e_{ji} \sim N(0, \sigma_e^2 c_{ji}^{-1})$), $\mathbf{x}_{ji}$ is the vector containing all observed predictor values for fixed effects for unit i in domain j, $\boldsymbol{\beta}$ is the vector of fixed model coefficients, c_{ji}^{-1} is the weight of the unit i in domain j in case of heteroskedasticity, v_j is the random area effect for domain j ($\upsilon_j \sim N(0, \sigma_\upsilon^2)$), and n_j is the sample size within domain j. In this study, potential heteroscedasticity was ignored, but a model where the observations in the target domain got a larger weight than those outside the domain was tested. In matrix form this is

$$(6) \quad \boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{v} + \mathbf{e}$$

where $\mathbf{Z}$ is an (N x J) indicator matrix of units belonging to domain j, $\boldsymbol{v}$ is the J x 1 matrix of all υ_j, and $\mathbf{X}$ and $\mathbf{e}$ are the matrices of all predictor values and random errors, respectively. This leads to a block-diagonal variance–covariance matric of the errors (Militino et al. 2007)

$$(7) \quad \boldsymbol{V} = \sigma_\upsilon^2 \mathbf{Z}\mathbf{Z}^{\mathbf{t}} + \sigma_e^2 \mathbf{C}^{-1}$$

where $\mathbf{C}$ is an ($N \times N$) diagonal weight matrix with elements c_{ji}, assuming in a constant correlation

$$(8) \quad \frac{\sigma_\upsilon^2}{\sigma_\upsilon^2 + \sigma_e^2}$$

within each domain j. With an alternative specification, it would also be possible to assume a continuous correlation depending on the distance between the units considered (Wadoux and Heuvelink 2023; Kangas et al. 2023).

The mean for domain j is (Militino et al. 2007; Mauro et al. 2017)

$$(9) \quad \mu_j = \bar{\boldsymbol{x}}_{\boldsymbol{j}}\boldsymbol{\beta} + v_j + \frac{1}{A_j}\sum_{i=1}^{N_j} e_{ji}$$

where $\bar{\boldsymbol{x}}_{\boldsymbol{j}}$ is the average vector of the values $\mathbf{x}_{\boldsymbol{ji}}$ in domain j, N_j is the total number of units within the domain j, v_j is the area (or domain) effect, and e_{ji} is the error related to the unit i within domain j. In 9, the area effect vector $\boldsymbol{v}$ is assumed to be estimable from data. However, if there are no observations from the domain (i.e., $n_j = 0$), the estimate of area-effect $\hat{v}_j = 0$. Moreover, if the N_j is large enough, the last term will be approximately zero and the MB estimator will simply be

$$(10) \quad \hat{\mu}_{\mathrm{MB},j} = \bar{\boldsymbol{x}}_{\boldsymbol{j}}\hat{\boldsymbol{\beta}}$$

A direct MB estimator is obtained, when the coefficients $\hat{\boldsymbol{\beta}}$ are estimated using only the observations within the domain j.

The (potentially) correlated errors need to be accounted for when estimating the variance (Magnussen and Breidenbach 2017). While the errors in formula (9) are not observed except for the units in the sample, their correlations can be estimated from the mixed-model correlation (8) or the semivariogram in the case of kriging. Then, the estimator based on the covariances between the residuals of predictions at different units (Breidenbach et al. 2016; Kotivuori et al. 2020) is

$$(11) \quad \mathrm{var}\left(\hat{\mu}_{\mathrm{MB},j}\right) = \bar{\boldsymbol{x}}_{\boldsymbol{j}}{}' \,\mathrm{var}\left(\hat{\boldsymbol{\beta}}\right) \bar{\boldsymbol{x}}_{\boldsymbol{j}} + \frac{1}{A_j^2}\sum_{i=1}^{N_j}\sum_{l=1}^{N_j}\widehat{\mathrm{cov}}\left(e_{ji}, e_{jl}\right)$$

That is, the variance of the mean prediction $\hat{\mu}_{\mathrm{MB},j}$ in the domain j is estimated from the estimated variances of the parameters and the estimated covariances (or variances in case $i = l$) of the errors of all N_j population elements or units within the domain. This can be fairly computing-intensive, especially if the correlations are assumed to vary as a function of distance.

When the MB estimator is calibrated using the observations from the domain j to estimate the area effect $\hat{v}_j$, the estimator of the mean is of the form (Mauro et al. 2017)

$$(12) \quad \hat{\mu}_{\mathrm{EBLUP},j} = \bar{\boldsymbol{x}}_{\boldsymbol{j}}\hat{\boldsymbol{\beta}} + \hat{v}_j$$

which can also be expressed as a composite estimator as

$$(13) \quad \hat{\mu}_j = f_j\bar{y}_{j,s} + (1 - f_j)\,\hat{\bar{y}}_{j,r}^{\mathrm{EBLUP}}$$

where $\bar{y}_{j,s}$ is the mean of the sample s and $\hat{\bar{y}}_{j,r}^{\mathrm{EBLUP}}$ is the EBLUP estimate for the nonsample units r and f_j is the sample ratio in domain j. Its MSE is estimated with (Militino et al. 2007, see also Rao and Molina 2015 chapter 5.2.6; Mauro et al. 2017; Breidenbach et al. 2018)

$$(14) \quad \mathrm{MSE}\left(\hat{\mu}_{\mathrm{EBLUP},j}\right) = g_{j,1} + g_{j,2} + 2g_{j,3} + g_{j,4}$$

where

$$(15) \quad g_{j,1} = (1 - \hat{\gamma}_j)\,\hat{\sigma}_\upsilon^2$$

$$(16) \quad g_{j,2} = \left(\bar{\mathbf{X}}_{j,r} - \hat{\gamma}_j\bar{\boldsymbol{x}}_j\right)^2 \mathrm{var}\left(\hat{\boldsymbol{\beta}}\right)$$

$$(17) \quad g_{j,3} = c_{j.}^{-2}\left(\hat{\sigma}_\upsilon^2 + \hat{\sigma}_e^2/c_{j.}\right)^{-3} \times \left[\hat{\sigma}_e^4\,\widehat{\mathrm{var}}\left(\hat{\sigma}_\upsilon^2\right) + \hat{\sigma}_\upsilon^4\,\widehat{\mathrm{var}}\left(\hat{\sigma}_e^2\right) - 2\hat{\sigma}_e^2\hat{\sigma}_{e\upsilon}^2\,\widehat{\mathrm{cov}}\left(\hat{\sigma}_e^2, \hat{\sigma}_\upsilon^2\right)\right]$$

$$(18) \quad g_{j,4} = \frac{\hat{\sigma}_e^2}{N_j^2}\sum_{i\in j_r} c_{ji}^{-1} = \frac{(N_j - n_j)\hat{\sigma}_e^2}{N_j^2}\bar{\mathbf{X}}_{j,r}$$

where $\hat{\sigma}_e^2$ and $\hat{\sigma}_\upsilon^2$ are estimators of the corresponding variances, $\gamma_j = \frac{\sigma_\upsilon^2}{\sigma_\upsilon^2 + \sigma_e^2/c_{j.}}$, and $c_{j.}$ is the sum of weights c_{ji} within domain j.

In addition to a random area-effect for adjusting the model locally, another option could be to localize the fixed part of the model. One way to localize the fixed part is to give more weight to the within-domain data. Lappi (2001) proposed a regression estimator or calibration estimator where the observations were weighted depending on the distance from the target region. In his proposal, the weight was decreasing as a function of the distance for distances shorter than a predefined range R, and zero for larger distances. The weighted estimators can be interpreted as a continuum between direct ($w = 0$, only observations from target domain are used) and indirect ($w = 1$, all observations are used) MB estimators.

3.3. Model-based estimators using K-NN

Utilizing the K-NN estimate either in a direct MA (Kangas et al. 2016) or indirect MB (McRoberts 2012) approach is possible. The K-NN estimator for one unit i in domain j is

$$(19)\quad \hat{y}_{ji} = \sum_{k=1}^{K} w_k^{ji} y_k^{ji}$$

where y_k^{ji} is measured value of variable y in the kth-nearest neighbor unit of i in domain j, and K is the number of the used neighbors. The weights w_k^{ji} are calculated based on the similarity between the target unit ji and its neighbors k, and this similarity is defined by a distance metric in predictor space, although some distance metrics also employ the values of a response variable of reference (sample) units (Packalén et al. 2012). In this study, we used the Euclidean distance as the metric.

When the estimator (19) is plugged to the difference estimator (3), a direct MA estimator is obtained. Yet, there is evidence that using K-NN as a model may lead to underestimating the variance (4) of the MA estimator (3) due to the nonlinearity of the K-NN predictor (Kangas et al. 2016), especially if the K-NN model is internal, i.e., estimated from the sample. In the MB framework, the estimator for the mean in domain j is the mean of the predictions

$$(20)\quad \hat{\mu}_{\text{KNN},j} = \frac{1}{A_j}\sum_{i=1}^{N_j} \hat{y}_{ji} = \frac{1}{A_j}\sum_{i=1}^{N_j}\sum_{k=1}^{K} w_k^{ji} y_k^{ji}$$

Here, we assume that the values of y_k^{ji} and $\hat{y}_{ji}$ are given per the area of the units. Thus, we divide the sum by area A_j in eq. 20.

This estimator can also be interpreted as an adjustment of the original sample weights based on the weights w each unit i gets in the formula over the domain j. It means that rather than summing up the K-NN model predictions for a given municipality, the results can be calculated as a weighted sum of the field plots, and instead of the sample weights, the weights are calculated by summing up the weights w_k^{ji} each field plot ji gets in the K-NN estimation (for details see Tomppo et al. 2011). However, suitable estimators for the variance and the MSE are largely missing.

McRoberts et al. (2007) and McRoberts (2012) proposed as a variance estimator

$$(21)\quad \text{var}\left(\hat{\mu}_{\text{KNN},j}\right) = \frac{1}{N_j^2}\sum_{i=1}^{N_j}\sum_{l=1}^{N_j}\widehat{\text{cov}}\left(\hat{\mu}_{ji}, \hat{\mu}_{jl}\right) = \frac{1}{N_j^2}\sum_{i=1}^{N_j}\widehat{\text{var}}\left(\hat{\mu}_{ji}\right) + 2\frac{1}{N_j^2}\sum_{i=1}^{N_j}\sum_{l>i}^{N_j}\widehat{\text{cov}}\left(\hat{\mu}_{ji}, \hat{\mu}_{jl}\right)$$

for the indirect estimate of mean in domain j. It is mimicking (11), but instead of the covariances between errors e, it is based on those between the pixel-level predictions. Since K-NN is nonparametric, parameter estimate errors are not involved. As the values in the original equation are assumed to be at per hectare level, the divisor in (21) is N_j rather than A_j as in this study. Like (11), also this formula can be computing intensive.

McRoberts et al. (2007) tested as unit-level variance estimate the variance calculated from the variation between the K-NN of unit i in domain j as

$$(22)\quad \widehat{\sigma_{ji}^2} = \sum_{k=1}^{K}\frac{\left(y_k^{ji} - \hat{y}_{ji}\right)^2}{K-1}$$

This does not necessarily reflect the actual estimation error at point ji, but rather the quality of the neighbors in general. They approximated the correlations needed for calculating the covariances for formula (21) using the proportion of common neighbors m from all neighbors K for each pair of units.

In this study, we tested both (21) and a modification of it. In the modified version, we used the K-NN approach to predict the error variance for each unit i within domain j as

$$(23)\quad \hat{e_{ji}^2} = \sum_{k=1}^{K} w_k^{ji}\left(e_k^{ji}\right)^2$$

That is, $\hat{e_{ji}^2}$ is the weighted average of the squared observed residuals of the K-NN of that unit. We approximated the between-unit covariances using a fixed correlation assumption with correlation (8) estimated from the area-effect model. These assumptions lead to a computationally quite tractable estimator

$$(24)\quad \widehat{\text{var}}\left(\hat{\mu}_{\text{KNN},j}\right) = \frac{1}{A_j^2}\sum_{i=1}^{N_j}\widehat{e_{ji}^2} + 2\frac{1}{A_j^2}\sum_{i=1}^{N_j}\sum_{l>i}^{N_j}\widehat{\text{cov}}\left(\hat{e}_{ji}, \hat{e}_{jl}\right) = \frac{1}{A_j^2}\left\{\left(\left(\sum_{i=1}^{N_j}\widehat{e_{ji}}\right)^2 - \sum_{i=1}^{N_j}\widehat{e_{ji}^2}\right)\cdot\hat{\rho} + \sum_{i=1}^{N_j}\widehat{e_{ji}^2}\right\}$$

where $\hat{e_{ji}} = \sqrt{\widehat{e_{ji}^2}}$. In preliminary tests the formula (21) gave systematically 6%–7% lower estimates than (24), given similar assumption on correlation. In addition to these estimators, bootstrap estimators have been proposed for estimating the variances and covariances for K-NN estimates.

3.4. Composites of design-based and model-based estimators

In some cases, it might be of interest to attempt to combine the best qualities of direct design-based and indirect MB estimators to a composite. Mixed model can as such be seen as a composite of an MB component and a design-based component (Lehtonen et al. 2003; Militino et al. 2007; eq. 12). However, if the MB estimator is based on K-NN or other nonparametric model, it may be of interest to use a composite of the form (Ghosh and Rao 1994):

$$(25)\quad \hat{\bar{y}}_c = w\hat{\bar{y}}_m + (1-w)\hat{\bar{y}}_d$$

where $\hat{\bar{y}}_d$ is the direct design-based estimate, $\hat{\bar{y}}_m$ is the indirect MB estimate, $\hat{\bar{y}}_c$ is the composite, and w is the weight.

Table 1. A summary of the tested methods.

Use of data outside target domain	Acronym	Model data	Model	Method	Formula for mean	Formula for variance
Direct	HT	–	–	Simple random sampling	(1)	(2)
	MA	Target domain	Linear	Model-assisted	(3)	(4)
	MB	Target domain	Linear	Model-based	(10)	(11)
	MALA	Large area	Linear	Model-assisted	(3)	(4)
Indirect	MBLA	Large area	Linear	Model-based	(10)	(11)
	MBW	Large area	Linear	Model-based with weights	(10)	(11)
	K-NN	Large area	K-NN	Model-based	(20)	(23)
Composite	COMP	Large area	Linear and K-NN	Composite of MA and K-NN	(24)	(28)
	MBEB	Large area	Linear	EBLUP	(12)	(14)
	KEB	Large area	Linear and K-NN	Composite of K-NN and EBLUP	(20 + 12)	(29)

Note: Three different linear and K-NN models were considered for each model-assisted and model-based estimator. HT, Horvitz–Thompson. MA, model-assisted; MB, model-based; MALA, model-assisted estimator for the large area; MBW, model-based with weight; MBLA, model-based estimator for the large area; K-NN, K-nearest neighbor; COMP, composite of MA and K-NN; MBEB, an MBLA estimator calibrated with EBLUP; KEB, combination of K-NN and EBLUP approach.

The optimal weight of the indirect estimator would be $w = \frac{\mathrm{MSE}\left(\hat{\bar{y}}_m\right)}{\left(\mathrm{MSE}\left(\hat{\bar{y}}_m\right)+\mathrm{var}\left(\hat{\bar{y}}_d\right)\right)}$, but the bias of the indirect estimator is generally unknown. Ghosh and Rao (1994) proposed as one option to estimate the MSE for the indirect MB estimator as

$$\text{(26)} \quad \mathrm{MSE}_{m,j} = \left(\hat{\bar{y}}_{m,j} - \hat{\bar{y}}_{d,j}\right)^2 - \mathrm{var}\left(\hat{\bar{y}}_{d,j}\right)$$

for each domain j. This leads to an estimator of optimal weight w as

$$\text{(27)} \quad w = \frac{\mathrm{MSE}\left(\hat{\bar{y}}_{m,j}\right)}{\left(\hat{\bar{y}}_{m,j} - \hat{\bar{y}}_{d,j}\right)^2}$$

This estimator (26) is, however, quite unstable, and led in our test simulations often to negative value of MSE and thus negative weight. Another weighting scheme based on averaged weight across the domains,

$$\text{(28)} \quad w = 1 - \frac{\sum_j \mathrm{var}\left(\hat{\bar{y}}_{d,j}\right)}{\sum_j \left(\hat{\bar{y}}_{m,j} - \hat{\bar{y}}_{d,j}\right)^2}$$

was assumed to be more stable (Ghosh and Rao 1994), and it was selected for this study. However, also this approach can produce negative weights in some cases, and therefore the weights were restricted to be between 0 and 1. The variance of the composite estimator is then the weighted sum of the variances of the mean estimates (assuming the indirect and direct estimators independent of each other) as

$$\text{(29)} \quad \mathrm{var}\left(\hat{\bar{y}}_c\right) = w^2 \mathrm{var}\left(\hat{\bar{y}}_m\right) + (1-w)^2 \mathrm{var}\left(\hat{\bar{y}}_d\right)$$

The assumption of independency may not be valid, but no suitable estimators exist for covariances of the used estimators.

Another possibility for making a composite is a combination of an indirect K-NN estimator and an EBLUP estimator using mixed model. As EBLUP estimator is already a composite of design-based and MB components (see eq. 13), so is its composite with K-NN estimator. The composite of these two estimators can be obtained by using the mixed model to predict the residuals of the K-NN estimate for each sampled unit i in domain j, and calculating the result as a sum of the K-NN mean (20) and estimate of the mean of the residuals for each domain j as a composite of a direct and EBLUP estimators (13) as

$$\text{(30)} \quad \hat{\mu} = \frac{1}{A_j}\sum_{i=1}^{N_j} \hat{y}_{ji} + f_j \bar{e}_{j,s} + \left(1 - f_j\right) \hat{\bar{e}}_{j,r}^{\mathrm{EBLUP}}$$

This approach enables us to use the EBLUP analytical formulas to calculate the variance of the estimator, and to utilize the within-area correlations estimable from the area effects. Bell et al. (2022) made a bias correction to K-NN small-area estimates by calculating as a design-weighted mean of the observed errors. The approach of combining K-NN and an area-effect model can be seen as a generalization of that approach. It was first used by Nothdurft et al. (2009).

3.5. Simulation test

To have spatially contiguous and mutually different small areas, the large area was divided to J domains with a k-means clustering approach with the coordinates as the sole predictor. We used values of $J = 3$, 5, 10, 20, 35, and 50 to introduce situations with the size of the small areas decreasing from an average of 77 273 population elements with $J = 3$ to an average of 4636 population elements with $J = 50$. These mean a range from about 120–1900 ha per domain.

We generated $S = 2000$ random samples s of size $n = 1000$ from the simulated test data. It means that with $J = 50$ the expected sample size within each domain was 20, so that direct estimation would be possible in all domains. The sampling fraction was 0.43%. The tested direct estimators were (Table 1): HT, MA, and MB within each domain (i.e., using only observations from the target domain for modelling), MA using the model estimated from the large-area sample including all domains (i.e., using all observations irrespective of the domain, model-assisted estimator for the large area (MALA)). The tested indirect estimators were MB estimators, which used a model-based estimator for the large area

Table 2. The bias, standard error, and RMSE by domain for the case with 10 domains and the best predictors.

	Domain	1	2	3	4	5	6	7	8	9	10
True mean ($m^3 \cdot ha^{-1}$)		88.55	57.21	93.1	85.93	140.46	103.03	101.78	54.36	84.43	65.26
	Method										
Bias ($m^3 \cdot ha^{-1}$)	HT	0.21	**0.01**	0.26	0.37	−0.37	**−0.07**	**−0.05**	0.16	−0.27	−0.08
	MA	−0.12	−0.12	−0.2	0.05	−0.13	−0.24	−0.25	**−0.1**	−0.24	−0.07
	MB	−0.12	−0.12	−0.2	0.05	−0.13	−0.24	−0.25	**−0.1**	−0.24	−0.07
	MALA	−0.03	−0.03	**−0.1**	**0.17**	**−0.1**	−0.11	−0.14	0.02	**0**	**0.03**
	MBLA	0.85	−0.68	*3.56*	*5*	*−7.16*	−2.21	0.66	*1.24*	*−4.94*	*1.61*
	MBW	1.2	−0.53	2.34	3.3	−4.63	−1.08	0.69	0.73	−3.6	1.05
	K-NN	*1.52*	*−0.96*	2.68	4.69	−4.17	*−5.26*	−0.68	0.6	−5.16	−0.05
	COMP	0.52	−0.41	1.03	1.91	−2.01	−2.3	−0.41	0.18	−2.18	−0.04
	MBEB	0.47	−0.44	1.8	2.63	−4.43	−1.15	0.3	0.63	−2.75	0.88
	KEB	**−0.05**	−0.13	2.24	3.07	−3.85	−1.83	0.1	0.66	−2.53	0.61
Standard error ($m^3 \cdot ha^{-1}$)	HT	*8.11*	*7.48*	*9.91*	*8.96*	*13.85*	*10.75*	*11.56*	*7.7*	*10.94*	*8.87*
	MA	3.33	2.43	3.69	3.43	6.2	3.91	4.54	2.67	3.85	3.35
	MB	3.33	2.43	3.69	3.43	6.2	3.91	4.54	2.67	3.85	3.35
	MALA	3.46	2.38	3.65	3.38	6.05	3.93	4.47	2.6	3.81	3.26
	MBLA	1.4	**0.74**	**1.42**	**1.29**	**3.07**	1.66	**1.94**	**0.91**	**1.21**	**0.91**
	MBW	1.57	0.87	1.77	1.6	4.1	2.26	2.69	1.09	1.64	1.19
	K-NN	**1.3**	0.88	1.48	1.47	3.56	**1.58**	1.99	1.12	1.31	0.95
	COMP	2.33	1.71	2.53	2.34	4.5	2.67	3.15	1.94	2.6	2.28
	MBEB	1.99	1.39	2.49	2.55	4.44	2.62	2.86	1.56	2.71	1.85
	KEB	2	1.39	2.58	2.66	4.58	2.78	2.96	1.68	2.5	1.77
RMSE ($m^3 \cdot ha^{-1}$)	HT	*8.12*	*7.48*	*9.92*	*8.96*	*13.85*	*10.75*	*11.56*	*7.7*	*10.95*	*8.87*
	MA	3.33	2.44	3.7	3.43	6.2	3.91	4.55	2.67	3.86	3.35
	MB	3.33	2.44	3.7	3.43	6.2	3.91	4.55	2.67	3.86	3.35
	MALA	3.46	2.38	3.65	3.39	6.05	3.93	4.47	2.6	3.81	3.26
	MBLA	**1.64**	**1.01**	3.84	5.16	7.79	2.77	**2.05**	1.53	5.08	1.85
	MBW	1.98	1.02	2.93	3.67	6.18	**2.51**	2.78	**1.31**	3.95	1.59
	K-NN	2	1.3	**3.06**	4.92	5.49	5.5	2.11	1.27	5.32	**0.96**
	COMP	2.39	1.76	2.73	**3.02**	**4.93**	3.52	3.17	1.94	3.39	2.28
	MBEB	2.05	1.46	3.07	3.66	6.27	2.86	2.88	1.68	3.86	2.05
	KEB	2	1.39	3.42	4.07	5.98	3.33	2.96	1.81	**3.56**	1.87

Note: The best result in bold, and worst result in red. See Table 1 for the methods. HT, Horvitz–Thompson; MA, model-assisted; MB, model-based; MALA, model-assisted estimator for the large area; MBW, model-based with weight; MBLA, model-based estimator for the large area; K-NN, K-nearest neighbor; COMP, composite of MA and K-NN; MBEB, an MBLA estimator calibrated with EBLUP; KEB, combination of K-NN and EBLUP approach.

(MBLA), a model that used all observations but with smaller weight (c_{ji}) for the observations outside the target domain (model-based with weight (MBW)), and K-NN. In addition, we tested three composite estimators, namely, an MBLA estimator calibrated with EBLUP (MBEB), a composite of MA and K-NN (COMP), and a combination of K-NN and EBLUP approach (KEB). The weights of observations in MBW were defined to be 1 for all observations within the area and 0.2 outside the area for most cases. In addition, we tested varying weights from 0 to 1.

From each sample *s*, three pre-selected linear models with varying accuracy were estimated and used in MA and MB estimators. The first of these models used the three best predictors found with leaps R package (See footnote 1). The best predictors for the linear model were Proportion of vegetation points relative to all points (%, first and last echo), and Inner volume. The RMSE of this model, calculated from all units, was 36.11 $m^3 \cdot ha^{-1}$ and R^2 0.87. The second model used three medium-level predictors, namely, Average intensity of first echoes, SumEntropy and Proportion of points above mean height. In this case, the RMSE was 68.5 $m^3 \cdot ha^{-1}$ and R^2 was 0.52. Finally, the third model used the three predictors having the smallest correlation with volume, namely, Average intensity of first echoes, Skewness of the vegetation point heights, and Proportion of points having cumulated at 20% of the height from all points. This model had RMSE 84.12 $m^3 \cdot ha^{-1}$ and R^2 was 0.28. The same three sets of variables were also used for K-NN estimation. Three different linear and K-NN models were used to have insights of the relative efficiency of the methods in different conditions. The varying quality can be interpreted as representing remote sensing materials with varying predictive power and varying populations where the existing models fit more or less adequately.

We estimated for each method the mean and standard deviation with the corresponding estimators. Then, we estimated

Fig. 1. The result of clustering in a case of 10 domains.

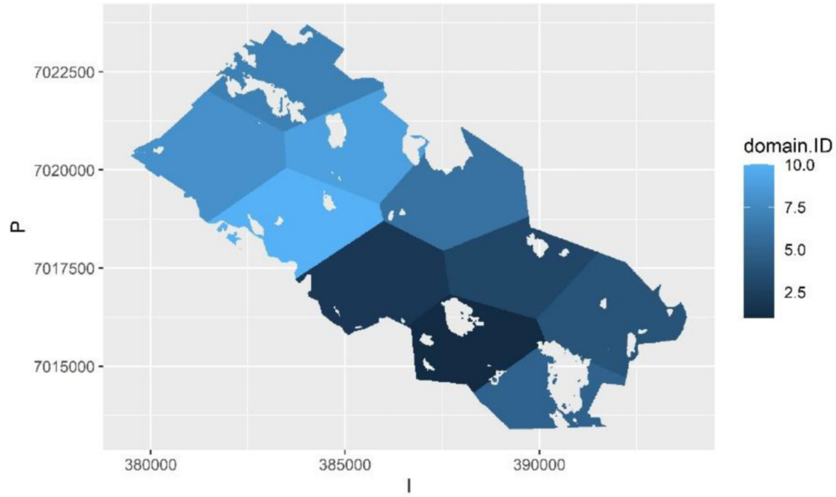


Table 3. The average empirical coverage of the confidence intervals for the three sets of predictors and two sets of area size ($J = 10$ and 50).

	J	Direct	Indirect	Composite
Best	10	0.93	0.61	0.90
Medium		0.94	0.60	0.91
Poor		0.94	0.45	0.92
Best	50	0.90	0.35	0.97
Medium		0.90	0.27	0.87
Poor		0.91	0.68	0.97

Fig. 2. The variation of domain-level RMSE of the different methods (see Table 1) for the case of 10 domains described with a violin plot. The black dot denotes the median RMSE among the domains. The models with the three best predictors were used (see the text for details). HT, Horvitz–Thompson; MA, model-assisted; MB, model-based; MBLA, model-based estimator for the large area; K-NN, K-nearest neighbor; MBW, model-based with weight; COMP, composite of MA and K-NN; MBEB, an MBLA estimator calibrated with EBLUP; KEB, combination of K-NN and EBLUP approach.

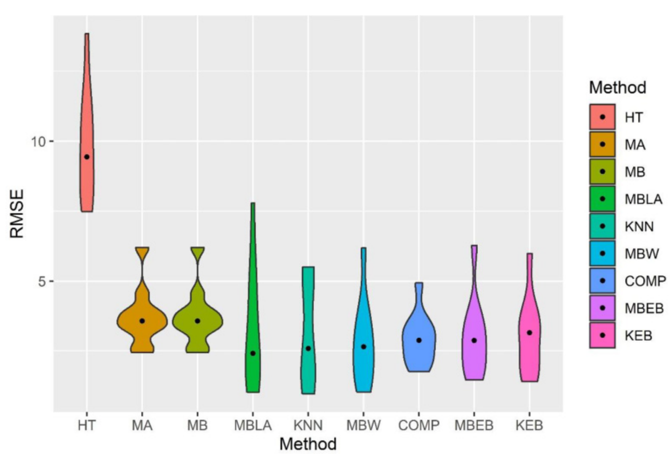
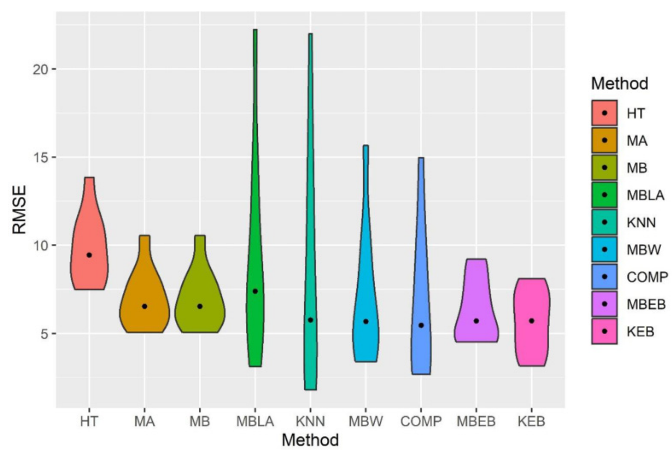


Fig. 3. The variation of domain-level RMSE of the different methods (see Table 1) for the case of 10 domains described with a violin plot. The black dot denotes the median RMSE among the domains. The models with the three medium-level predictors were used (see the text for details). HT, Horvitz–Thompson; MA, model-assisted; MB, model-based; MBLA, model-based estimator for the large area; K-NN, K-nearest neighbor; MBW, model-based with weight; COMP, composite of MA and K-NN; MBEB, an MBLA estimator calibrated with EBLUP; KEB, combination of K-NN and EBLUP approach.



the bias as a difference between the mean of mean estimates $\hat{\bar{y}}_{sj}$ for domain j from samples $s = 1, \ldots, S$ and true mean μ_j, i.e.,

$$\text{bias} = \sum_{s=1}^{S} \hat{\bar{y}}_{sj} - \mu_j \tag{31}$$

Fig. 4. The variation of domain-level RMSE of the different methods (see Table 1) for the case of 10 domains described with a violin plot. The models with the three poorest predictors were used (see the text for details). The black dot denotes the median RMSE among the domains. HT, Horvitz–Thompson; MA, model-assisted; MB, model-based; MBLA, model-based estimator for the large area; K-NN, K-nearest neighbor; MBW, model-based with weight; COMP, composite of MA and K-NN; MBEB, an MBLA estimator calibrated with EBLUP; KEB, combination of K-NN and EBLUP approach.

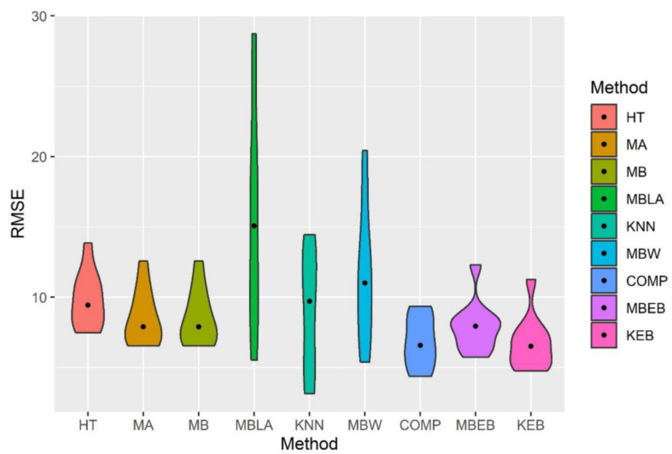
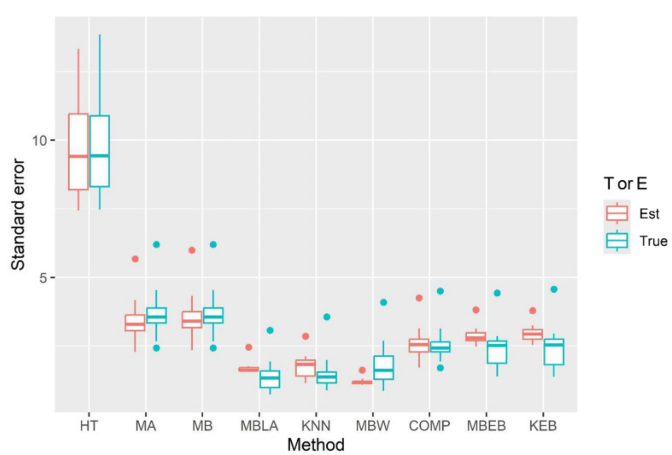


Fig. 5. The box-plot of domain-level true (estimated from the between sample variation eq. 29) and estimated (mean of the estimated obtained using the variance estimators listed in Table 1) standard errors for a case of 10 domains with the 3 best predictors. HT, Horvitz–Thompson; MA, model-assisted; MB, model-based; MBLA, model-based estimator for the large area; K-NN, K-nearest neighbor; MBW, model-based with weight; COMP, composite of MA and K-NN; MBEB, an MBLA estimator calibrated with EBLUP; KEB, combination of K-NN and EBLUP approach.



the true standard deviation as the standard deviation between the mean estimates of the S simulations, i.e.,

$$(32)\quad \text{std} = \sqrt{\frac{1}{S-1}\sum_{s=1}^{S}\left(\hat{\bar{y}}_{sj} - \widehat{\bar{\bar{y}}}_{sj}\right)^2}$$

and the true RMSE as

$$(33)\quad \text{RMSE} = \sqrt{\frac{1}{S-1}\sum_{s=1}^{S}\left(\hat{\bar{y}}_{sj} - \mu_j\right)^2}$$

We also calculated for each of the estimators the empirical coverage of the confidence interval (CI). That is, for each domain, we counted the proportion of samples s where the true mean was included to the estimated CI

$$(33)\quad \text{CI} = \widehat{\bar{y}}_{sj} \pm 1.96\,\widehat{\sigma}\left(\widehat{\bar{y}}_{sj}\right)$$

We used the R-package JoSAE (Breidenbach et al. 2018) for calculating the EBLUP variances and the R package yaImpute (Crookston and Finley 2007) for calculating the K-NN means and variances.

4. Results

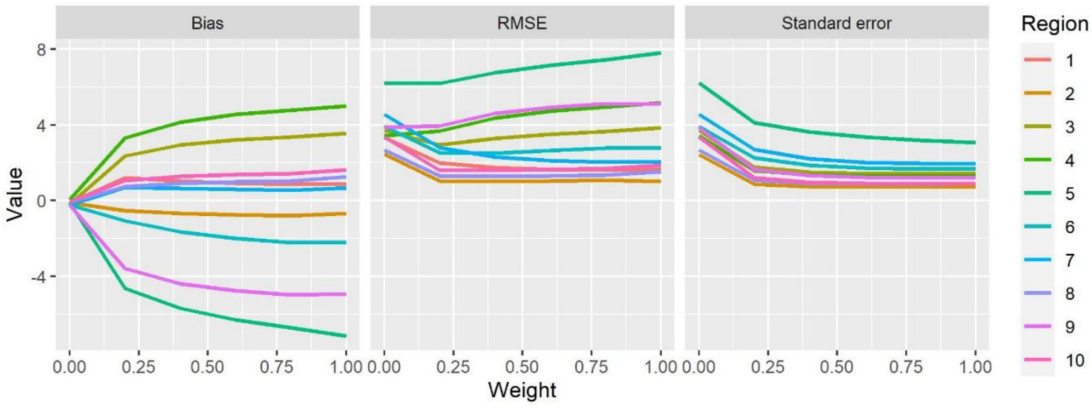
With all domain sizes, the HT estimator was by far the least accurate when the MA and MB estimators used the best predictors. With the three best predictors, the estimated models were so accurate that using HT even as part of the composite estimator with K-NN (not included to the tables) produced poorer results than all MB and MA estimators. Therefore, the composite estimator COMP was chosen to be the composite of the MB K-NN and design-based MA estimator for all three sets of predictors.

The true standard errors, biases, and RMSEs for the best model are presented in Table 2. As the simulated domains are not of interest in themselves, domain-level results are presented only for the case with 10 domains, with an expected sample size of 100 (Fig. 1). Unless otherwise stated, the results with varying J were pooled together.

All indirect MB methods borrowing strength from other domains (MBLA, MBW, and K-NN, see Table 1) had high bias for at least some of the domains (Table 2). This means that either there were differences between the domains that were not explained with the predictors used in the model (but possibly could have been with other predictors), or the coefficients should have varied among the domains (i.e., model should have been localized at least to some extent). This can also be seen from the empirical coverage of the CIs: for the indirect estimators the average empirical coverage was 61% with the best three predictors (Table 3). The variation between the domains was high; the empirical coverage of, for example, MBLA varied from 8% to 99% between the domains. This means that irrespective of the sample, the best model could not depict the differences between the small areas. The EBLUP calibration (MBEB, KEB) and composite estimator COMP were able to remove part of the biases, but not all. For the composite methods, the average coverage was 90%.

Even though the bias was high at some domains, the indirect methods borrowing strength from the neighboring areas

Fig. 6. The dependence of true bias, RMSE, and standard error on the weight given to observations outside the target domain with model-based with weight method in the case of 10 domains and the best three predictors.



(MBLA, MBW, and K-NN) had the smallest average RMSEs in 7 domains out of 10 (Table 2). The composites COMP, MBEB, and KEB showed a compromise result, and either COMP or KEB had the smallest RMSE in 3 domains out of 10 (Table 2). The direct model-assisted (MA) and model-based (MB) based only on data from the domain of interest have higher median RMSE than the indirect and composite methods (Fig. 2). Surprisingly, the composite of K-NN and MA, COMP, had shortest tails in the distribution, i.e., lowest probability for very poor and very good results.

When the same calculations were carried out with the medium-level model, the indirect methods borrowing strength from outside the domain of interest proved to be vulnerable to the model deficiencies. The bias for all these methods was very high at least for some of the areas (Fig. 3; Appendix A). In this case, the indirect MBLA was the poorest regarding to the RMSE in two domains, K-NN in three domains but HT was still the worst in 5 domains out of 10. With the medium model, the empirical coverage of the indirect methods was 60% (Table 3), and the variation between domains for MBLA was from 0% to 99%. The composite KEB of K-NN and EBLUP was the best in 3 domains, and indirect K-NN was the best in 5 domains, and the weighted model (MBW) in 1 domain out of 10. The empirical coverage of composite estimators was 91%. It is notable that even with a poorer model, the direct methods were the best only in 1 domain out of 10, and even in that domain best direct estimator was MALA, which used the large-area model rather than MA, which used the domain-level model.

With the medium-level predictors, the direct MA and MB as well as the EBLUP estimators (MBEB, KEB) were more robust than the indirect estimators, but still, in terms of median level of RMSE, K-NN, and MBW perform as well as the composites (Fig. 3). In fact, the domain-level RMSE of the HT method was in all domains greater than the median RMSE of all other methods.

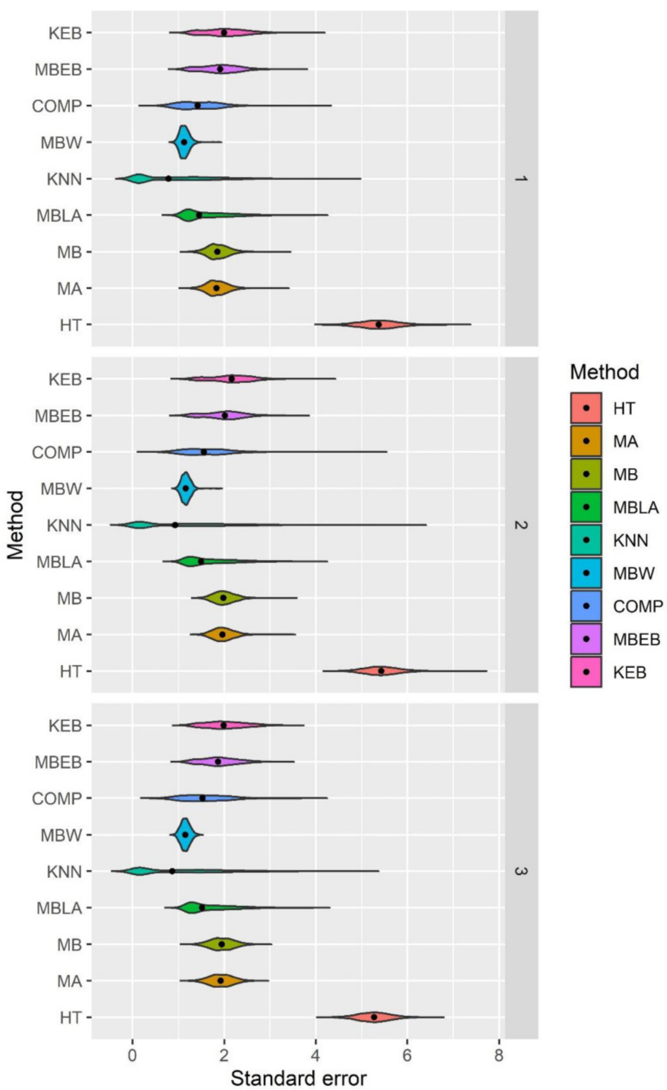
With the poorest three predictors, the HT was still the worst regarding the standard error in all cases, and regarding the RMSE in three domains out of 10 (Appendix B). With respect to the bias, the indirect MB estimates based on the large-area model were worst in all 10 domains, and with respect to the RMSE in 6 domains out of 10. The empirical coverage of indirect estimators CI reduced to 45% (Table 3). It is notable that with the more local models (K-NN local in predictors space and MBW localized with the weights), the indirect estimators were far more robust than the model-based method based on the large-area model. When the model was poor, the composite estimators (COMP, KEB, and MBEB) were best in 7 domains out of 10 (Appendix B; Fig. 4).

The problem with K-NN is that there is no proper estimator for the standard deviation or RMSE. Here the standard error estimator (24) is based on smoothed squared residuals and an average correlation for the whole area. Yet, the estimator provided reasonably good estimates, overestimating the true standard error to some extent (Fig. 5). This also applies to the case with poor predictors. The worst estimates of standard error were obtained for the MB estimator with weighted data, MBW, where the estimated standard error was a gross underestimate of the true standard error. All the indirect estimators ignore the bias, and therefore underestimating the standard deviation is particularly problematic. The RMSE estimates obtained with estimator (14) for EBLUP (MBEB) and (29) for the composite of K-NN and EBLUP (KEB) were also slight overestimates compared to the true standard error, but these estimators also acknowledge the bias.

When the weights of the indirect model-based method (MBW) were varied from 0 (direct MB) to 1 (indirect MBLA), some trends could be noted. When the weight increased, also the absolute value of domain-level bias increased, in the beginning fast but later slowly (Fig. 6). Instead, the true standard errors decreased, first fast and then slowly. Thus, there was a clear trade-off between precision and bias. Based on RMSE the optimum weight was 0.2 for most of the 10 domains.

The results above are averages over the 2000 repetitions of the sampling for the 10 areas. The selected small-area estimators should, however, work in a robust fashion for all samples. For this, we analyzed the distribution of the sample-level results. It is notable that while all the methods produced poor results for the regions for some of the samples, the HT estimator was clearly the poorest method with the three best

Fig. 7. The distributions of the estimated standard errors from the 2000 samples in the case of three domains and the best three predictors. HT, Horvitz–Thompson; MA, model-assisted; MB, model-based; MBLA, model-based estimator for the large area; K-NN, K-nearest neighbor; MBW, model-based with weight; COMP, composite of MA and K-NN; MBEB, an MBLA estimator calibrated with EBLUP; KEB, combination of K-NN and EBLUP approach.



predictors, and weighted MB estimator borrowing strength from neighboring areas (MBW) produced most robust results in terms of the standard error (Fig. 7). While the mean and median of standard error by domain with K-NN was in line or smaller than with other methods, the method resulted in quite high standard error estimates occasionally. The composite estimators and direct MB and MA were fairly similar with respect to the variation.

A robust method should produce competitive (wrt other methods) true RMSE with varying sizes of domains and varying quality models. Furthermore, since the bias cannot be estimated from a sample, for all MB estimators, and composite estimators involving an MB component, the accuracy measured with the available formulas should give a sufficient estimate compared to the true accuracy, i.e., RMSE. To summarize the results, we divided the areas to three classes (small < 10 000 population elements, 10 000 < medium < 30 000, and larger > 30 000 population elements). When the results were summarized with respect to the true RMSE, using a model proved to be a good choice in all cases (Fig. 8). It turned out that with a good model, indirect approach was competitive, but composites were overall the best performing alternatives. Direct approaches were competitive only in the largest areas. When the results were summarized with respect to the absolute value of bias, the indirect methods were the least competitive option (Fig. 9). Due to the large biases, the standard error estimators are grossly underestimating the true accuracy for these indirect methods. For the EBLUP methods, the bias was smaller, but also the availability of formula for MSE is important in obtaining a good estimate of accuracy. The importance of bias is also reflected in the empirical coverage of the CIs, which plunged to 27%–68% with the $J = 50$ but remained at 90%–92% for direct methods and 87%–97% for the composite methods (Table 3).

When the absolute value of bias was modelled as a function of three factors, namely, the domain size class, model quality, and the estimation approach, it turned out that use of indirect estimator had the largest effect on the bias (Table 4), poor model was the second most important and small size of the domain the third. This model explained 39% of the variation in the estimated biases.

When the number of domains increases, and they get smaller, the probability of not sampling any or enough plots from all domains increases. Then, the usefulness of direct methods not borrowing strength from neighboring areas diminishes (Fig. 10; Table 1). With 50 domains indirect MB methods borrowing strength from neighboring areas are more robust than direct methods based only on the domain-level observations. With 50 domains and the poorest model, 12 samples out of all 2000 samples, and across all domains did not produce a value, meaning it was not possible to calculate an estimate at all due to lack of observations. For even smaller domains correspondent to the stand-level, i.e., for domains of 1–4 ha, the proportion would obviously be much higher.

5. Discussion

While calculating estimates is possible in a case study with one sample, assessing the properties of the estimators requires either independent validation data (e.g., Breidenbach et al. 2016) or simulated wall-to-wall data (e.g., Magnussen and Breidenbach 2017). The benefit of independent validation data are that it includes the true complexity of data, but, on the other hand, it is costly and typically also includes sampling errors. The simulated data may be overly simple, and while it enables examining the properties of different estimators in detail, the structure of the test data may affect to the results. In this study, we attempted to create as realistic wall-to-wall data as possible, to get useful insights of the various estimators. To assess if the results were due to this specific population, we simulated also other populations. The results obtained from these other populations were practically iden-

Fig. 8. Box-plot of the true RMSE as a function of the approach, model quality, and domain size.

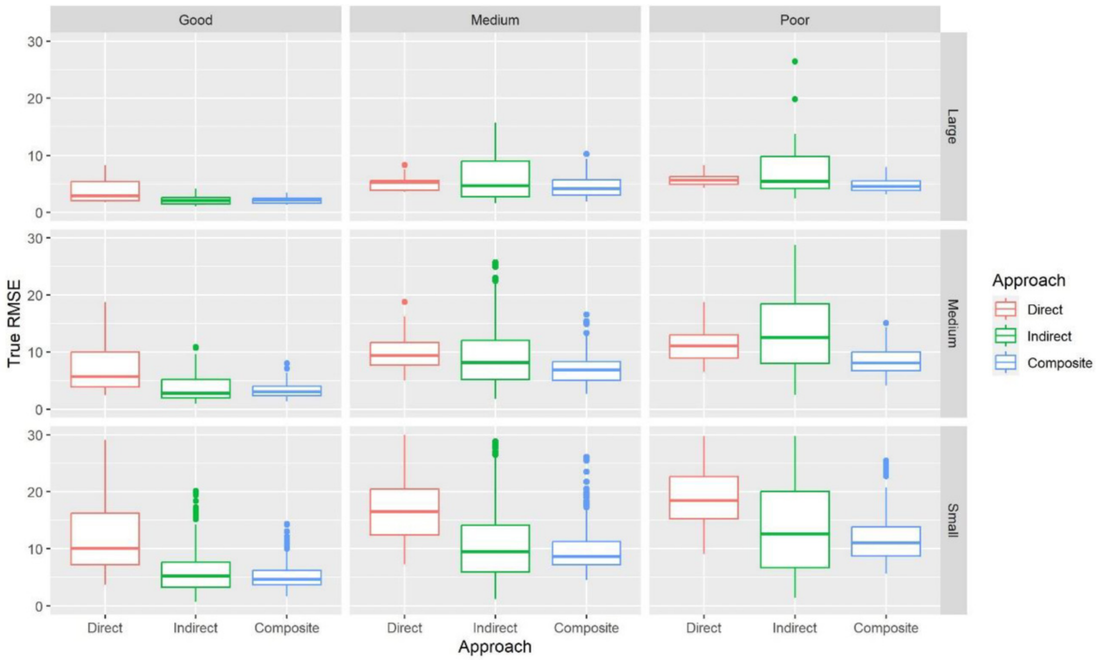
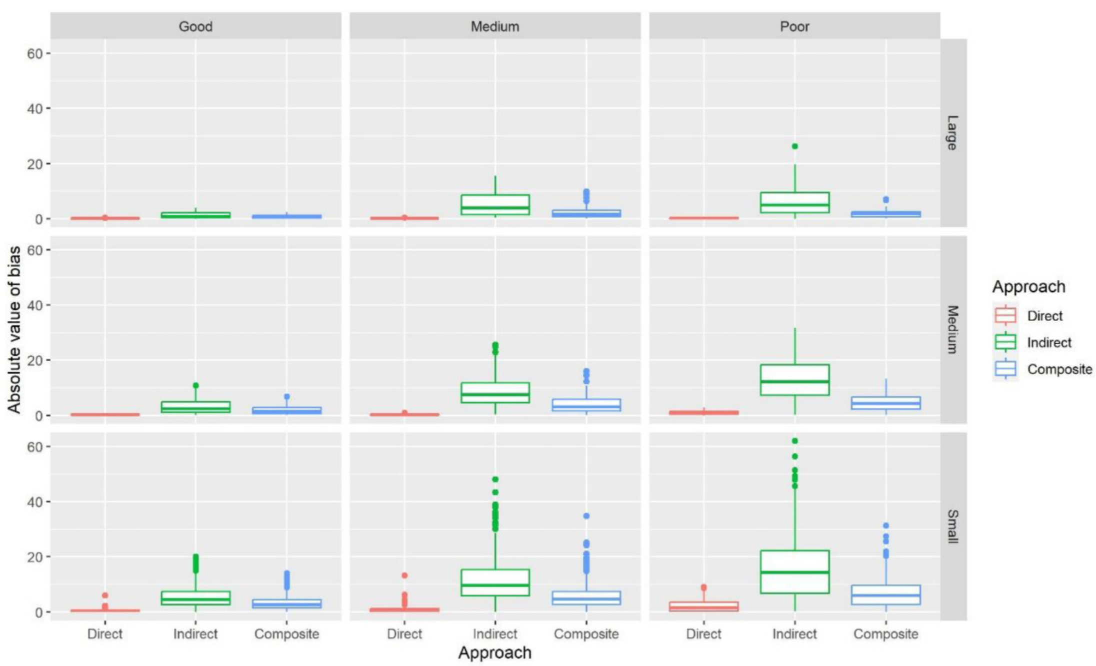


Fig. 9. Box-plot of the bias as a function of the approach, model quality, and domain size.



tical with the results obtained with this population, and thus the results from these populations are not shown.

The small-area estimators should be utilizable in all kinds of small areas, they should be robust against the poor models and poor assumptions, and they should have a reliable estimator of accuracy. None of the methods tested in this study fulfill all these requirements.

Our results imply that models and auxiliary information are highly beneficial in small-area estimation, even when the quality of the model is far from perfect. Likewise, using data from outside the target domain is also beneficial. Moreover, it can be argued that for at least part of the users of the data the difference between a systematic bias and a random error is not relevant, the total error is what is important. Thus, the method producing the smallest RMSE would be the optimal one. From this point of view, none of the direct methods were optimal in this study, except for the largest size class of domains, meaning domains larger than circa 700 ha or 30 000

Table 4. Coefficients of the factors Model (good, medium, poor), Size (large, medium, small), and Approach (direct, indirect, composite) for a linear model predicting the absolute value of bias.

	Estimate	Std. error	*t* value
Intercept	−4.736	0.414	−11.429
Medium model	2.985	0.231	12.947
Poor model	4.814	0.231	20.879
Medium domain size	2.037	0.417	4.886
Small domain size	3.827	0.386	9.918
Indirect estimator	9.058	0.231	39.282
Composite estimator	3.773	0.231	16.361

Note: Large domain size, direct estimator, and a good model is the reference case.

units. The design-based HT estimator was often the poorest estimator with respect to RMSE even when the MA and MB methods used a very poor model. Similar conclusion on the usefulness of models compared to a SRS (or HT estimator) was also made by Breidenbach and Astrup (2012) and Frescino et al. (2022).

A very popular special case of MA estimation is PS, where the only auxiliary variable is the group indicator. Since these methods require quite a large sample from each domain to work, their use is restricted to fairly large "small areas" (Myllymäki et al. 2017; Pulkkinen et al. 2018; Haakana et al. 2019, 2020; White et al. 2021; Frescino et al. 2022; Finlay et al. 2024). Yet, these studies also confirm that it is always beneficial to utilize a model for the estimation, if such a model is available.

It should be noted that our whole study area (5900 ha) is much smaller than the areas considered as "small areas" in many other studies. For instance, Bell et al. (2022) had small areas starting from 68 000 to 250 ha and used as a smallest sampling fraction 2%. Thus, the interpretation of what constitutes a small area, for which estimates are required, is relative to the size of the study area and to the sampling fraction, i.e., to the amount of the sample plots available for estimation within each domain.

The selection between a direct and indirect approach can be based on the trade-off between the bias and the standard error. When the model is good one and is expected to describe the differences between the domains well, the bias component may not have a high importance, but the poorer the model, the more important the bias component is. Thus, when good predictors are available, it may make sense to use indirect methods, but the poorer the model is assumed to be, the more useful it is to use composite or direct methods. Unfortunately, unambiguous definition of the required accuracy for a "good enough" model is not possible, as it depends on the variation across domains (which may further be dependent of the sizes and the heterogeneity of the domains). However, our results suggest that if the coefficient of determination of the model is around 0.85–0.95, it might be sufficient for using an indirect approach rather than a composite estimator. Even then, it is evident that the uncertainty of the results is seriously underestimated at some areas due to the potential bias.

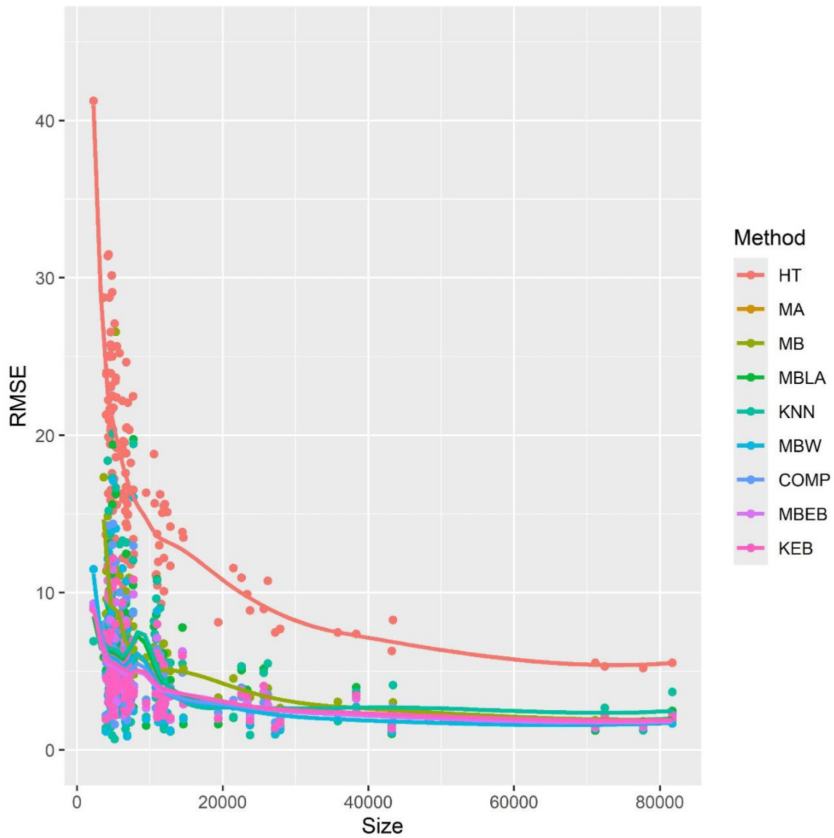
The indirect MB methods borrowing the strength from neighboring areas, MBLA and MBW, produced really good results with a good model, but with the poorer models the results deteriorated markedly. For instance, the bias in the worst domain more than tripled from the best model to the worst model (7.16 → 15.96 → 25.75) in the case of 10 domains. Both MBW and K-NN can be interpreted as somewhat local estimators (i.e., giving more weight to nearby observations either in real or in feature space), and both performed better than the basic large-area model MBLA when the medium or poor predictors were used. Based on these results, localizing the fixed part of the model is recommendable. However, usefulness of localization, e.g., in terms of restricting the range where the nearest neighbors should be selected, or how they should be weighted, remains to be studied in the future. Likewise, localizing the regression model with random coefficients for the fixed parameters might also be a good option, especially as an analytical formula for the variance exists. When the model coefficient of determination decreases, the desirability of composite methods increases.

It is also possible to control the trade-off by reducing the weight given to the observations outside of the area and give smaller weight in cases when the differences between the small areas cannot be accurately reflected. In this study, we only showed the results of fixed weights: simulating distance-based weights for domains in the shorter edges of the large area would have been quite restricted. Based on the results, it seems likely that the optimal range in distance-based weighting depends on the quality of the predictors, but that remains to be confirmed in future studies.

All the composite estimators, namely EBLUP estimator (MBEB), the COMP, and the composite of K-NN and EBLUP (KEB), provided fairly accurate estimates for both the domain-level means and their accuracy across various conditions. It means that the bias of the composite estimates was low compared to the indirect estimators and their uncertainty was not underestimated. A little surprisingly, the EBLUP estimators (MBEB and KEB) were the best (i.e., had best accuracy among the estimators) only in one of the 10 domains for the good quality model, while COMP was the best in two domains. With the medium-level model, KEB was the best in 3 domains out of 10, and with the worst model one of the composite estimates was the best in 7 domains out of 10.

The usefulness of the EBLUP approach has also been acknowledged in many previous studies (e.g., Breidenbach et al. 2016; Magnussen and Breidenbach 2017; Frescino et al. 2022). Good results are not, however, self-evident. For instance, Breidenbach et al. (2016) note that it is important to acknowledge the heteroskedasticity of the residuals to get a good estimate for the uncertainty. In this study, the (potential) heteroscedasticity was ignored, which may have resulted underestimation of the uncertainty. Magnussen and Breidenbach (2017) go as far as stating that more than one observation should be available from all domains, and the results of the empirical coverages of the CIs in this study confirm that

Fig. 10. The true RMSE of the methods as a function of the domain size (number of units within the domain) using the three best predictors. Each dot corresponds to the RMSE (30) of a domain computed from 2000 replicated random samples. The line is estimated with loess using a span of 0.5. HT, Horvitz–Thompson; MA, model-assisted; MB, model-based; MBLA, model-based estimator for the large area; K-NN, K-nearest neighbor; MBW, model-based with weight; COMP, composite of MA and K-NN; MBEB, an MBLA estimator calibrated with EBLUP; KEB, combination of K-NN and EBLUP approach.



even a very good model may introduce bias to one or more of the domains. It remains to be studied if more balanced sampling methods than SRS can reduce this tendency.

The K-NN and MA composite COMP was the best more often (4 cases) than the EBLUP estimators (MBEB 1, KEB 2 cases) with this poorest model. It can be concluded that importance of localization of the fixed part increases when the quality of the predictors of the model decrease: in the case of K-NN, such localization can be achieved with neighbors selected from nearby, both in real space and in feature space. For weighted models, it means giving more weight to observations from nearby. However, the EBLUP estimators were the most robust in terms of not overestimating the true accuracy, and not producing very poor RMSEs often. So, results of our simulation study imply that in cases where the good quality of the model cannot be verified, selecting an EBLUP or a composite estimator is the wisest choice. However, the realized results in any real case depend on how large differences there are between the areas, and how large portion of them can be explained with the model. It is likely that the variations between the domains would be greater if the study area were larger and the domains were smaller (i.e., more varying conditions). Moreover, the proportion of within-domain variance of the total variance, i.e., the homogeneity of the domains may have an effect. It is also likely that in a simulated data set the differences are smaller than in real forest data. Unfortunately, obtaining a more realistic data for testing small-area estimators is difficult to obtain.

The composite of K-NN and EBLUP, KEB, has the attractive possibility of being able to combine the K-NN method and an analytical estimator of variance. It provided a good estimator for the true RMSE with all three models sets in a case of 10 domains. Our estimator (24) for the variance of K-NN estimator showed that the covariance of the neighboring unit errors has a larger effect on the domain variance estimate than the actual pixel-level variances. While the within-area correlations of observed errors were very small, the covariances still had a tenfold effect on the variance estimate for a given area: assuming no correlation could produce a standard error of 0.25, while a 0.02 correlation produced a standard error of 2.5. It means that the combination of K-NN and EBLUP is recommendable, unless accurate estimator of the covariances for formula (24) is found.

Conclusions

The composites in the form of EBLUP (MBEB) and the composite of the K-NN and MA (COMP) and the K-NN and EBLUP (KEB) are the most recommendable of the tested estimators, whenever there is one or more observations from the target domain. The possibility of combining a flexible model in K-NN and the analytical error estimator also looks attractive.

Localizing the models by weighting the nearby observations more than those further away can also be recommended, but the optimal way to localize the models is not clear. Unfortunately, the bias and thus the true accuracy for K-NN as well as for all indirect methods remains unattainable. The variance estimator for K-NN based on smoothed squares of the residuals performed fairly well as an estimator for precision, but it should be further investigated under different setups. Future studies should explore computationally efficient means for estimating within-domain covariances for the errors of K-NNs' estimates. Indirect estimators are unavoidable in cases no observations are available from all the domains. However, it is evident that the uncertainty of MB estimators is underestimated, and even with a very good model the empirical coverage of the CI can be near zero, i.e., the CI does not necessarily include the true values for all small areas. The potential for unknown bias needs therefore always be clarified to the users of the data.

Article information

History dates

Received: 15 March 2024
Accepted: 8 October 2024
Accepted manuscript online: 22 October 2024
Version of record online: 31 January 2025

Copyright

Data availability

All codes and the modelling data are available from the corresponding author with request.

Author information

Author ORCIDs

Annika Kangas https://orcid.org/0000-0002-8637-5668

Author contributions

Formal analysis: AK
Funding acquisition: AK, MM, PP
Methodology: AK, MM, PP
Writing – original draft: AK
Writing – review & editing: MM, PP

Competing interests

The authors declare there are no competing interests.

Funding information

The study was supported by the Research Council of Finland through the project "Is climate smart forestry a utopia if the preferences of landowners are not considered? (UTOPIA)" under Grant 352782, the European Union Horizon Europe (HORIZON) Research & Innovation programme under the Grant Agreement No. 101056907, and the Research Council of Finland's flagship "Forest–Human–Machine Interplay—Building Resilience, Redefining Value Networks, and Enabling Meaningful Experiences" (UNITE) (Grant number 357909).

References

Astrup, R., Rahlf, J., Bjørkelo, K., Debella-Gilo, M., Gjertsen, A.K., and Breidenbach, J. 2019. Forest information at multiple scales: development, evaluation and application of the Norwegian forest resources map SR16. Scand. J. For. Res. **34**(6): 484–496. doi:10.1080/02827581.2019.1588989.

Balasz, A., Liski, E., Tuominen, S., and Kangas, A. 2022. Comparison of neural networks and K-nearest neighbors methods in forest stand variable estimation using airborne laser data. ISPRS Open J. Photogramm. Remote Sens.

Bell, D.M., Wilson, B.T., Werstak, C.E., Jr, Oswalt, C.M., and Perry, C.H. 2022. Examining *K*-nearest neighbor small area estimation across scales using National Forest Inventory data. Front. For. Global Change, **5**: 763422. doi:10.3389/ffgc.2022.763422.

Breidenbach, J., and Astrup, R. 2012. Small area estimation of forest attributes in the Norwegian National Forest Inventory. Eur. J. For. Res. **131**: 1255–1267. doi:10.1007/s10342-012-0596-7.

Breidenbach, J., Magnussen, S., Rahlf, J., and Astrup, R. 2018. Unit-level and area-level small area estimation under heteroscedasticity using digital aerial photogrammetry data. Remote Sens. Environ. **212**: 199–211. doi:10.1016/j.rse.2018.04.028.

Breidenbach, J., McRoberts, R.E., and Astrup, R. 2016. Empirical coverage of model-based variance estimators for remote sensing assisted estimation of stand-level timber volume. Remote Sens. Environ. **173**: 274–281. doi:10.1016/j.rse.2015.07.026. PMID: 28148972.

Crookston, N.L., and Finley, A.O. 2007. yaImpute: an R package for KNN imputation. J. Stat. Software, **23**(10). ISSN 1548-7660. doi:10.18637/jss.v023.i10.

Finley, A.O., Andersen, H-.E., Babcock, Cook, B.D., Morton, D.C., and Banerjee, S. 2024. Models to support forest inventory and small area estimation using sparsely sampled LiDAR: a case study involving G-LiHT LiDAR in Tanana, Alaska. J. Agric. Biol. Environ. Stat. 695–722. doi:10.1007/s13253-024-00611-3.

Frescino, T.S, McConville, K.S., White, G.W., Toney, J.C., and Moisen, G.G. 2022. Small area estimates for national applications: a database to dashboard strategy using FIESTA. Front. For. Global Change, **5**: 779446. doi:10.3389/ffgc.2022.779446.

Ghosh, M., and Rao, J.N.K. 1994. Small-area estimation: an appraisal. Stat. Sci. **9**(1): 55–93.

Haakana, H., Heikkinen, J., and Kangas, A. 2019. Efficiency of post-stratification for a large-scale forest inventory—case Finnish NFI. Ann. For. Sci. **76**: 9. doi:10.1007/s13595-018-0795-6.

Haakana, H., Katila, M., Heikkinen, J., and Kangas, A. 2020. Precision of exogenous post-stratification in small area estimation based on a continuous national forest inventory. Can. J. For. Res. **50**(999): 359–370. doi:10.1139/cjfr-2019-0139.

Haralick, R.M., Shanmugam, K., and Dinsteein, J. 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybern. **SMC-3**: 610–621. doi:10.1109/TSMC.1973.4309314.

Kangas, A. 1996. Small area estimates using model based methods. Can. J. For. Res. **26**: 758–766. doi:10.1139/x26-085.

Kangas, A., Myllymäki, M., Gobakken, T., and Næsset, E. 2016. Model-assisted forest inventory with parametric, semi-parametric and non-parametric models. Can. J. For. Res. **46**: 855–868. doi:10.1139/cjfr-2015-0504.

Kangas, A., Myllymäki, M., and Mehtätalo, L. 2023. Understanding uncertainty in forest resources maps. Silva. Fennica. **57**: 22026.

Kotivuori, E., Kukkonen, M., Mehtätalo, L., Maltamo, M., Korhonen, L., and Packalen, P. 2020. Forest inventories for small areas using drone imagery without in-situ field measurements. Remote Sens. Environ. **237**: 111404. doi:10.1016/j.rse.2019.111404.

Lappi, J. 2001. Forest inventory of small areas combining the calibration estimator and a spatial model. Can. J. For. Res. **31**: 1551–1560. doi:10.1139/x01-078.

Lehtonen, R., Särndal, C.E., and Veijanen, A. 2003. The effect of model choice in estimation for domains, including small domains. Surv. Methodol. **29**(1): 33–44.

Magnussen, S., and Breidenbach, J. 2017. Model-dependent forest stand-level inference with and without estimates of stand-effects. Forestry, **90**: 675–685. doi:10.1093/forestry/cpx023.

Mäkisara, K., Katila, M., and Peräsaari, J. 2022. The multi-source National Forest Inventory of Finland—methods and results 2017 and 2019. Natural resources and bioeconomy studies 90/2022.

Mauro, F., Monleon, V.J., Temesgen, H., and Ford, K.R. 2017. Analysis of area level and unit level models for small area estimation in forest inventories assisted with LiDAR auxiliary information. PLoS One, **12**(12): e0189401. doi:10.1371/journal.pone.0189401. PMID: 29216290.

McRoberts, R.E. 2012. Estimating forest attribute parameters for small areas using nearest neighbors techniques. For. Ecol. Manage. **272**: 3–12. doi:10.1016/j.foreco.2011.06.039.

McRoberts, R.E., Tomppo, E., Finley, A., and Heikkinen, J. 2007. Estimating areal means and variances of forest attributes using the K-nearest neighbors technique and satellite imagery. Remote Sens. Environ. **111**(4): 466–480. doi:10.1016/j.rse.2007.04.002.

Militino, A.F., Ugarte, M.D., and Goicoa, T. 2007. A BLUP synthetic versus an EBLUP estimator: an empirical study of a small area estimation problem. J. Appl. Stat. **34**(2): 153–165. doi:10.1080/02664760600994893.

Myllymäki, M., Gobakken, T., Næsset, E., and Kangas, A. 2017. The efficiency of post-stratification compared to model-assisted estimation. Can. J. For. Res. **47**: 515–526. doi:10.1139/cjfr-2016-0383.

Nothdurft, A., Saborowski, J., and Breidenbach, J. 2009. Spatial prediction of forest stand variables. Eur. J. For. Res. **128**: 241–251. doi:10.1007/s10342-009-0260-z.

Packalén, P., Temesgen, H., and Maltamo, M. 2012. Variable selection strategies for nearest neighbor imputation methods used in remote sensing-based forest inventory. Can. J. Remote Sens. **38**(5): 557–569. doi:10.5589/m12-046.

Pulkkinen, M., Ginzler, C., Traub, B., and Lanz, A. 2018. Stereo-imagery-based post-stratification by regression-tree modelling in Swiss National Forest Inventory. Remote Sens. Environ. **213**: 182–194. doi:10.1016/j.rse.2018.04.052.

Rao, J.N.K., and Molina, I. 2015. Small area estimation. 2nd ed. John Wiley & Sons, Inc. doi:10.1002/9781118735855.

Särndal, C.E., Swensson, B., and Wretman, J. 1992. Model assisted survey sampling. Springer-Verlag. 694p.

Tomppo, E., Heikkinen, J., Henttonen, H., Ihalainen, A., Katila, M., Mäkelä, H., et al. 2011. Designing and conducting a forest inventory-case: 9th National Forest Inventory of Finland. Springer Verlag.

Tuominen, S., Pitkänen, T., Balázs, A., and Kangas, A. 2017. Improving multi-source National Forest Inventory by 3D aerial imaging. Silva Fenn. **51**(4). doi:10.14214/sf.7743.

Véga, C., Renaud, J.-P., Durrieu, S., and Bouvier, M. 2016. On the interest of penetration depth, canopy area and volume metrics to improve Lidar-based models of forest parameters. Remote Sens. Environ. **175**: 32–42. doi:10.1016/j.rse.2015.12.039.

Wadoux, A.C., and Heuvelink, G.B.M. 2023. Uncertainty of spatial averages and totals of natural resource maps. Methods Ecol. Evol. **14**: 1320–1332. doi:10.1111/2041-210X.14106.

White, G.W., McConville, K.S., Moisen, G.G., and Frescino, T.S. 2021. Hierarchical Bayesian small area estimation using weakly informative priors in ecologically homogeneous areas of the interior western forests. Front. For. Global Change, **4**: 752911. doi:10.3389/ffgc.2021.752911.

Appendix A

Table A1. The bias, standard error and RMSE by domain for a case with 10 areas and three medium-level predictors.

	Domain	1	2	3	4	5	6	7	8	9	10
Bias ($m^3 \cdot ha^{-1}$)	HT	0.21	**0.01**	0.26	**0.37**	−0.37	−0.07	**−0.05**	0.16	−0.27	−0.08
	MA	**0.01**	−0.04	−0.16	**0.37**	−0.72	**−0.03**	−0.47	−0.21	−0.14	−0.09
	MB	**0.01**	−0.04	−0.16	**0.37**	−0.72	**−0.03**	−0.47	−0.21	−0.14	−0.09
	MALA	0.13	0.08	**0**	**0.37**	**−0.04**	0.05	−0.11	**−0.11**	**−0.03**	**−0.02**
	MBLA	*22.05*	*7.46*	*−2.07*	4.62	*−11.92*	−14.89	−9.1	*6.6*	*−6.71*	*3.25*
	MBW	15.38	4.86	−1.71	3.04	−7.89	−10.37	−6.72	4.39	−5.09	2.3
	K-NN	21.84	3.15	1.97	*11.13*	−2.52	*−15.96*	*−11.81*	2.65	−6.49	1.06
	COMP	14.63	2.08	1.33	7.64	−1.93	−10.77	−8.21	1.71	−4.4	0.71
	MBEB	7.46	2.01	−0.67	1.47	−4.55	−3.98	−2.84	1.57	−2.08	0.88
	KEB	6.02	0.53	0.39	2.72	−0.85	−3.56	−3.19	0.37	−1.69	0.1
Standard error ($m^3 \cdot ha^{-1}$)	HT	*8.11*	*7.48*	*9.91*	*8.96*	*13.85*	*10.75*	*11.56*	*7.7*	*10.94*	*8.87*
	MA	5.55	5.05	6.73	6.32	10.53	7.41	8.35	5.45	7.95	6.19
	MB	5.55	5.05	6.73	6.32	10.53	7.41	8.35	5.45	7.95	6.19
	MALA	5.52	5.05	6.66	6.27	10.46	7.4	8.25	5.48	7.85	6.12
	MBLA	2.82	2.14	2.32	2.31	**3.39**	2.27	2.39	1.93	2.07	2.01
	MBW	2.89	2.42	2.94	2.93	4.44	3.27	3.35	2.39	3.01	2.5
	K-NN	**2.61**	**1.31**	**2.18**	**2.27**	4.08	**1.98**	**2.14**	**1.27**	**1.74**	**1.45**
	COMP	3.18	2.13	3.03	3.09	5.14	2.94	3.27	2.31	3.19	2.58
	MBEB	5.39	4.04	4.98	4.9	7.66	6.39	6.42	4.36	5.96	4.55
	KEB	5.42	3.12	4.73	5.14	7.67	5.96	6.25	3.57	5.36	3.97
RMSE ($m^3 \cdot ha^{-1}$)	HT	8.12	7.48	*9.92*	8.96	*13.85*	10.75	11.56	*7.7*	*10.95*	*8.87*
	MA	5.55	5.05	6.73	6.33	10.55	7.41	8.36	5.45	7.95	6.19
	MB	5.55	5.05	6.73	6.33	10.55	7.41	8.36	5.45	7.95	6.19
	MALA	**5.52**	5.05	6.66	6.28	10.46	7.4	8.25	5.48	7.85	6.12
	MBLA	*22.23*	*7.76*	3.11	5.16	12.4	15.06	9.41	6.88	7.02	3.82
	MBW	15.65	5.43	3.4	**4.23**	9.05	10.87	7.51	5	5.92	3.4
	K-NN	22	3.41	**2.94**	*11.36*	**4.8**	*16.08*	*12*	**2.94**	6.72	**1.8**
	COMP	14.97	**2.98**	3.31	8.24	5.49	11.16	8.83	2.88	5.44	2.67
	MBEB	9.2	4.51	5.02	5.11	8.9	7.53	**7.02**	4.63	6.31	4.64
	KEB	8.1	3.16	4.75	5.81	7.72	**6.95**	**7.02**	3.59	**5.62**	3.97

Note: The best result in bold, and worst result in red. HT, Horvitz–Thompson; MA, model-assisted; MB, model-based; MALA, model-assisted estimator for the large area; MBW, model-based with weight; MBLA, model-based estimator for the large area; K-NN, K-nearest neighbor; COMP, composite of MA and K-NN; MBEB, an MBLA estimator calibrated with EBLUP; KEB, combination of K-NN and EBLUP approach.

Appendix B

Table B1. The bias, standard error and RMSE by domain for a case with 10 areas and the three poorest predictors.

	Domain	1	2	3	4	5	6	7	8	9	10
True mean ($m^3 \cdot ha^{-1}$)		88.55	57.21	93.1	85.93	140.46	103.03	101.78	54.36	84.43	65.26
Bias ($m^3 \cdot ha^{-1}$)	HT	0.21	**0.01**	0.26	**0.37**	−0.37	−0.07	**−0.05**	0.16	−0.27	−0.08
	MA	0.66	0.61	0.41	0.97	1.73	0.71	1.49	0.33	0.81	0.71
	MB	0.66	0.61	0.41	0.97	1.73	0.71	1.49	0.33	0.81	0.71
	MALA	**0.1**	0.08	**0.14**	0.55	**−0.08**	**0.1**	0.19	**0.03**	**−0.11**	**0.07**
	MBLA	9.45	28.59	−11.98	−4.96	−25.75	−17.72	−10.4	18.31	−6.11	18.37
	MBW	6.46	17.29	−9.79	−4.38	−19.78	−13.2	−8.28	11.33	−5.34	12.13
	K-NN	14.17	8.9	−3.36	2.05	−12.78	−11.55	−6	10.21	−3.73	12.91
	COMP	8.07	5.15	−1.66	1.65	−6.83	−6.12	−2.77	5.77	−1.65	7.55
	MBEB	2.09	4.69	−2.11	−0.43	−6.63	−2.88	−1.87	2.9	−1.28	3.34
	KEB	5.13	2.74	−1.63	0.47	−6.88	−4.43	−2.96	3.01	−1.96	4.16
Standard error ($m^3 \cdot ha^{-1}$)	HT	8.11	7.48	9.91	8.96	13.85	10.75	11.56	7.7	10.94	8.87
	MA	6.51	6.71	7.68	7.26	12.45	8.82	10.56	6.7	9.86	8.09
	MB	6.51	6.71	7.68	7.26	12.45	8.82	10.56	6.7	9.86	8.09
	MALA	6.45	6.8	7.78	7.2	12.12	8.99	10.26	6.69	9.67	7.99
	MBLA	3.03	2.9	2.48	2.47	**3.96**	2.63	2.94	2.49	2.5	2.75
	MBW	3.1	3.34	3.17	3.13	5.15	3.87	4.22	3.1	3.65	3.5
	K-NN	**2.79**	**1.77**	**2.42**	**2.37**	4.75	**2.52**	**2.91**	**1.83**	**2.19**	**2.17**
	COMP	3.86	3.63	4.16	4.06	6.4	4.67	5.37	3.72	5.08	4.25
	MBEB	5.34	6.59	6.66	6.14	10.37	7.96	8.67	6.06	8.07	7.05
	KEB	5.16	4.27	5.06	4.74	8.92	6.72	6.55	4.57	5.9	5.39
RMSE ($m^3 \cdot ha^{-1}$)	HT	8.12	7.48	9.92	8.96	13.85	10.75	11.56	7.7	10.95	8.87
	MA	6.54	6.74	7.69	7.33	12.57	8.85	10.66	6.71	9.9	8.12
	MB	6.54	6.74	7.69	7.33	12.57	8.85	10.66	6.71	9.9	8.12
	MALA	6.45	6.8	7.78	7.22	12.12	8.99	10.26	6.69	9.67	7.99
	MBLA	9.92	28.74	12.23	5.54	26.06	17.91	10.8	18.48	6.6	18.58
	MBW	7.17	17.61	10.29	5.39	20.44	13.76	9.29	11.74	6.47	12.62
	K-NN	14.44	9.08	**4.14**	**3.14**	13.64	11.82	6.67	10.37	**4.33**	13.09
	COMP	8.94	**6.3**	4.48	4.38	**9.36**	**7.7**	**6.04**	6.86	5.34	8.67
	MBEB	**5.74**	8.08	6.98	6.16	12.31	8.46	8.87	6.72	8.17	7.8
	KEB	7.27	5.07	5.31	4.76	11.26	8.04	7.19	**5.47**	6.22	**6.81**

Note: The best result in bold, and worst result in red. HT, Horvitz–Thompson; MA, model-assisted; MB, model-based; MALA, model-assisted estimator for the large area; MBW, model-based with weight; MBLA, model-based estimator for the large area; K-NN, K-nearest neighbor; COMP, composite of MA and K-NN; MBEB, an MBLA estimator calibrated with EBLUP; KEB, combination of K-NN and EBLUP approach.