NRC Research Press

ARTICLE

National forest inventories in the service of small area estimation of stem volume

Steen Magnussen, Daniel Mandallaz, Johannes Breidenbach, Adrian Lanz, and Christian Ginzler

Abstract: This study introduces five facets that can improve inference in small area estimation (SAE) problems: (1) model groups, (2) test of area effects, (3) conditional EBLUPs, (4) model selection, and (5) model averaging. Two contrasting case studies with data from the Swiss and Norwegian national forest inventories demonstrate the five facets. The target variable of interest was mean stem volume per hectare on forested land in 108 Swiss forest districts (FD) and in 14 Norwegian municipalities (KOM) in the County of Vestfold. Auxiliary variables from airborne laser scanning (Switzerland) and photogrammetric point clouds (Vestfold) with full coverage and a resolution of 25 m × 25 m (Switzerland) and 16 m × 16 m (Vestfold) were available. Only the data metric mean canopy height was statistically significant. Ten linear fixed-effects models and three mixed linear models were assessed. Area effects were statistically significant in the Swiss case but not in Vestfold case. A model selection based on AIC favored separate linear regression models for each FD and a single common regression model in Vestfold. Model averaging increased, on average, an estimated variance by 15%. Reported estimates of uncertainty were consistently larger than corresponding unconditional EBLUPs.

Key words: model groups, tests of area effect, model selection, model averaging, conditional EBLUP.

Résumé : Cette étude présente cinq aspects qui peuvent améliorer l'inférence dans les cas d'estimations pour de petites régions géographiques: (1) les groupes de modèles; (2) les tests de l'effet de région; (3) les meilleures prédictions empiriques linéaires sans biais (EBLUP) conditionnelles; (4) la sélection de modèles; et (5) la combinaison de modèles. Deux études de cas contrastantes, utilisant des données des inventaires forestiers nationaux de la Suisse et de la Norvège sont utilisées pour démontrer ces cinq aspects. La variable d'intérêt ciblée était le volume moyen à l'hectare par tige des terrains forestiers de 108 Districts forestiers (DF) de la Suisse et de 14 municipalités du Comté de Vestfold en Norvège. Des variables auxiliaires provenant de couvertures complètes de balayage laser aéroporté (Suisse) et de nuages de points photogrammétriques (Vestfold), avec une résolution de 25 m × 25 m (Suisse) et de 16 m × 16 m (Vestfold) étaient disponibles. La hauteur moyenne du couvert forestier était la seule donnée métrique statistiquement significative. Dix modèles linéaires à effets fixes et trois modèles linéaires mixtes ont été évalués. Les effets de région étaient statistiquement significatifs dans le cas de la Suisse mais pas dans le cas du Vestfold. Une sélection de modèle sur la base du critère d'information d'Akaike a préféré des modèles de régression linéaires séparés pour chaque DF, mais un seul modèle de régression commun pour le Vestfold. La combinaison de modèles a augmenté la variance estimée de 15 % en moyenne. Les estimations d'incertitude rapportées étaient toujours plus grandes que les meilleures prédictions empiriques linéaires sans biais non conditionnelles correspondantes. [Traduit par la Rédaction]

Mots-clés : groupes de modèles, tests de l'effet de région, sélection de modèles, combinaison de modèles, meilleure prédiction empirique linéaire sans biais (EBLUP) conditionnelle.

Introduction

National forest inventories (NFIs) with sample plots established on a national grid provide accurate national and regional (state-wide) estimates of important forest resource variables (McRoberts et al. 2010). In a context of sustainable forest management and stewardship issues, the demands from forest industry and local administrations for forest resource information at smaller spatial scales have grown considerably (Corona et al. 2002; Gillespie 1999; Ståhl et al. 2011; Tomppo et al. 2010).

Through the use of auxiliary variables (**X**) correlated with the study variable (*y*), and small area estimation (SAE) methods (e.g., Rao 2003), a NFI may also meet more local information needs (Breidenbach and Astrup 2012; Goerndt et al. 2013; Pfeffermann 2013). Forestry SAE applications are numerous (Breidenbach and Astrup 2012; Goerndt et al. 2011; McRoberts 2011; Steinmann et al. 2013).

A design-based SAE (Rao 2003, p. 10) derived exclusively from data collected in the NFI plots within an area of interest (AOI) may not meet a desired level of accuracy. Fortunately, advances and improvements in remote sensing technologies and data analysis have created opportunities for NFIs to improve the accuracy of SAEs (Gallaun et al. 2010; Kangas 1996; Maltamo et al. 2009; Tomppo et al. 2008). Auxiliary variables from medium- to high-resolution satellites, airborne scanners, and aerial photography dominate in forestry applications (Goerndt et al. 2013; Katila and Tomppo 2006).

The information about *y* contained in **X** can be exploited in one of two ways: (*i*) area-level models (Rao 2003, ch. 7.1) or (*ii*) unit-level models (Rao 2003, ch. 7.2). Summary statistics of **X** (total, mean, variance) for each AOI are used as predictors in area-level models.

Received 30 October 2013. Accepted 9 May 2014.

S. Magnussen. Natural Resources Canada, Canadian Forest Service, 506 West Burnside Road, Victoria, BC V8Z 1M5, Canada.
D. Mandallaz. Chair of Land Use Engineering, ETH Zurich, CH 8092 Zurich, Switzerland.
J. Breidenbach. Norwegian Forest and Landscape Institute, Postboks 115, 1431 Ås, Norway.
A. Lanz and C. Ginzler. Swiss Federal Research Institute, WSL, Zürcherstrasse 111, 8903 Birmensdorf ZH, Switzerland.
Corresponding author: Steen Magnussen (e-mail: steen.magnussen@nrcan.gc.ca).

Published at www.nrcresearchpress.com/cjfr on 12 May 2014.

In a unit-level model, the relationship between y and $\mathbf{X}$ is formulated at the level of a single NFI plot. The sampling frames (populations) for the two sets of AOIs are the forested land in the union of the AOIs. The survey variable is, in both cases, stem volume per ha of forested land. Population and AOI-level totals (means) of the auxiliary variable(s) are computed based on a complete tessellation of the populations into N equal sized units (a post-sampling construct).

The improved accuracy of SAEs obtained with an area-level or a unit-level model comes from what is known as *borrowing strength*. From Lehtonen et al. (2003) we have: "Borrowing strength is generally understood to mean that the estimator in use depends on data on the variable of interest, denoted y, from related areas, or more generally from a larger area, in an effort to improve the accuracy for the small area". Dependencies arise through model-defined relationships between y and one or more auxiliary variables shared across two or more AOIs. The means (totals) of the auxiliary variables are known for the AOIs and, in many cases, known also for each unit in an AOI (Longford 2005, p. 214; Olsen et al. 1999; Rao 2003, p. 20). The models used for borrowing strength can be parametric, semi-parametric, or non-parametric (Goerndt et al. 2011; McRoberts 2011; Opsomer et al. 2008; Saei and Chambers 2003). Models with a mixture of fixed and random effects are popular (Breidt 2004; Lohr and Prasad 2003), but estimates of variance components can be imprecise in SAE applications with few AOIs (e.g., <20) and when sample sizes vary greatly among AOIs (Datta and Lahiri 2000; Fuller 2009, p. 314; Jiang and Lahiri 2006*a*; Longford 2005, p. 241–246).

Statistical estimators applicable to a wide range of sampling designs and model choices have been developed for SAE problems (Chambers and Clark 2012, p. 161; Datta 2009; Jiang et al. 2011; Longford 2005, p. 207). Geographically weighted regression (GWR) is a more recent example (Salvati et al. 2012). In GWR, borrowing strength relies on the presence of a positive distance dependent spatial correlation.

Agencies adhering to a design-based (DB) inference paradigm choose model-assisted (MA) estimators for SAE problems (Lehtonen and Veijanen 2009; Särndal et al. 1992, p. 386). The model-dependent (MD) approach to SAE offers an alternative (Chambers and Clark 2012, p. 161; Lehtonen et al. 2005; Lehtonen and Veijanen 2009). An MA estimator is design-consistent and nearly design-unbiased, but it can be imprecise. A MD estimator is design-biased, but accuracy can be good if the model fits well in an AOI. Bias will be serious if the selected model is inadequate. Sample sizes in an AOI may, however, be too small to allow a meaningful assessment of model-fit and precision (Longford 2005, p. 195; Rao 2003, p. 110).

Recently, Mandallaz (2013) proposed several flexible infinite population MA SAE regression estimators of variance formulated as a sum of sampling errors and errors in estimated model parameters. Särndal (2007) gives an overview of finite population MA estimators of variance, such as the generalized regression estimator, the calibrated estimator, and the model-calibrated estimator by Wu (2003). Deville (1999) provides technical details on variance estimation. Additional details on variance estimation for complex models can be found in, for example, Jiang and Lahiri (2006*a*), Rao (2003), Särndal et al. (1992), and Lehtonen and Veijanen (2009).

All regression estimators of variance for an AOI estimate of a total or a mean are functions of the estimated regression residuals. An analyst in pursuit of a minimum residual variance may therefore be tempted to favor a more complex model over a simpler model despite the risk of an optimistic estimate of variance (Hastie et al. 2005, p. 200). The limited appeal of borrowing strength in MA SAE espoused by Estevao and Särndal (2004) may, in part, reflect this phenomenon.

This study introduces five facets of SAE that can improve inference in SAE problems: (1) model groups (Lehtonen and Veijanen 2009), (2) tests of area effects (Datta et al. 2011), (3) conditional EBLUP (Chambers and Clark 2012, p. 177), (4) model selection (Shen et al. 2004), and (5) model averaging (Claeskens and Hjort 2008).

(1) A model group is a subset of AOIs with an anticipated similar relationship between y and $\mathbf{X}$. Sample data from members of a model group are pooled for the purpose of model estimation. Hence, borrowing strength is limited to "similar" AOIs which, ceteris paribus, support simpler and more parsimonious models. A single AOI can be member of more than one model group.

(2) Testing of area effects affords model simplification and improved precision of SAEs when the test supports the absence of area effects.

(3) Conditional EBLUPs provide a measure of uncertainty conditional on the actual values of the estimated area effects in AOIs. In contrast, the popular unconditional EBLUP gives a measure of uncertainty that is averaged over the area effects (Jiang and Lahiri 2006*b*).

(4) A model selection based on Akaike's information criterion (AIC) protects against optimistic estimates of uncertainty by penalizing model complexity (over fitting).

(5) Model averaging provides SAE estimates with an improved trade-off between variance and bias. When no single model has a much larger AIC than the runner up model(s), model averaging provides a more effective protection against optimistic estimates of uncertainty than selecting the single "best" model (Chatfield 1995; Draper 1995).

We demonstrate the five facets in a context of unit-level SAE of mean stem volume per hectare in 108 forest districts (FD) in Switzerland and 14 municipalities (KOM) in the Norwegian County of Vestfold. In the Swiss case, the number of sample plots and AOIs are fairly large and suggests a design-based direct MA estimator for each AOI. In the Norwegian case the AOIs are much smaller, and sample sizes across AOIs range from 1 to 35, which a priori suggests that an indirect estimator with a single common regression model would be appropriate.

The Swiss data indicated a significant spatial correlation of random FD effects. As a complement to the unit-level analyses, we briefly sketch spatial EBLUP results (Chandra et al. 2007; Pratesi and Salvati 2008).

Plot-specific values of stem volume per hectare were provided by the Swiss and Norwegian national forest inventories. Auxiliary data came from either a country-wide airborne laser scanning (Switzerland) or in the form of photogrammetric point clouds covering the AOIs (Vestfold).

Material and Methods

Study areas

Small area estimation of mean stem volume (y) per hectare of forested land is carried out for two populations: (*i*) the forested area in108 Swiss Forstkreise (henceforth called forest districts or simply FDs) and (*ii*) the forested area in 14 Norwegian municipalities (kommuner, KOM) in the County of Vestfold, southwest of Oslo. Estimation is based on a combination of national forest inventory data and auxiliary data from airborne laser scanning (Switzerland) and photogrammetric point clouds (Vestfold). The auxiliary data constitute a census (full coverage) of the 108 FDs and the 14 KOMs.

Swiss data

Field data from the third (2004–2006) Swiss national forest inventory (SNFI) were used. Plots in the SNFI are located on a regular $\sqrt{2}$ km × $\sqrt{2}$ km national grid (Brändli 2010). Thus the sampling design is systematic without stratification, but field data were only collected in locations considered as forest.

For each SNFI plot with a center located in a forest, the stem volume (m^3) per ha was computed and used as the study variable (y). The volume was calculated from measurements of stem diameters of live trees at a reference height of 1.3 m (DBH) and application of localized tariff models (Kaufmann 1999). All trees with a

Table 1. Unit-level linear models and estimates of Akaike's information criterion ($\widehat{\mathrm{AIC}}$, $\widehat{\mathrm{cAIC}}$) and "model probability" (w_r).

	Swiss forest districts	$\hat{\mathrm{AIC}}$ ($w_r \times 10^3$)	Vestfold municipalities	$\hat{\mathrm{AIC}}$ ($w_r \times 10^3$)	Min n_i =
1	$y_{ij} = \beta_0 + \beta_1 x_{ij} + e_{ij}$	68857 (4)	$y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + e_{ij}$	**1625 (354)**	1
2	$y_{ij} = \beta_{0i} + \beta_1 x_{ij} + e_{ij}$	68787 (6)	$y_{ij} = \beta_{0i} + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + e_{ij}$	1638 (185)	2
3	$y_{ij,G1\ni i} = \beta_0^{G1} + \beta_1 x_{ij} + e_{ij}$	68258 (7)	$y_{ij,G1\ni i} = \beta_0^{G1} + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + e_{ij}$	1664 (50)	1
4	$y_{ij,G1\ni i} = \beta_0 + \beta_1^{G1} x_{ij} + e_{ij}$	69325 (0)	$y_{ij,G1\ni i} = \beta_0 + \beta_1^{G1} x_{ij} + \beta_2 x_{ij}^2 + e_{ij}$	1687 (16)	1
5	$y_{ij,G2\ni i} = \beta_0^{G2} + \beta_1 x_{ij} + e_{ij}$	68129 (138)	$y_{ij,G2\ni i} = \beta_0^{G2} + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + e_{ij}$	1672 (16)	1
6	$y_{ij,G2\ni i} = \beta_0 + \beta_1^{G2} x_{ij} + e_{ij}$	68585 (16)	$y_{ij,G2\ni i} = \beta_0 + \beta_1^{G2} x_{ij} + \beta_2 x_{ij}^2 + e_{ij}$	1660 (34)	1
7	$y_{ij,G3\ni i} = \beta_0^{G3} + \beta_1^{G3} x_{ij} + e_{ij}$	68509 (23)	n.a.		1
8	$y_{ij,G3\ni i} = \beta_0 + \beta_1^{G3} x_{ij} + e_{ij}$	68318 (57)	n.a.		1
9	$y_{ij,G3\ni i} = \beta_{0i} + \beta_1^{G3} x_{ij} + e_{ij}$	68113 (149)	n.a.		2
10	$y_{ij} = \beta_{0i} + \beta_{1i} x_{ij} + e_{ij}$	**67996 (258)**	$y_{ij} = \beta_{0i} + \beta_{1i} x_{ij} + \beta_{2i} x_{ij}^2 + e_{ij}$	1719 (3)	4
11	$y_{ij} = \mu + u_{0i} + e_{ij}$	68339 (51)	$y_{ij} = \mu + u_{0i} + e_{ij}$	1768 (0)	2
12	$y_{ij} = \beta_0 + u_{0i} + \beta_1 x_{ij} + e_{ij}$	68222 (89)	$y_{ij} = \beta_0 + u_{0i} + \beta_1 x_{ij} + \beta_2 x_{ij}^2 + e_{ij}$	1642 (151)	2
13	$y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i}) x_{ij} + e_{ij}$	68135 (134)	$y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i}) x_{ij} + (\beta_2 + u_{2i}) x_{ij}^2 + e_{ij}$	1643 (144)	4

Note: y_{ij} is stem volume ($m^3 \cdot ha^{-1}$) in the jth plot in the ith AOI, x_{ij} is the mean canopy height (ch, unit: m) in the jth plot in the ith AOI; β_p, p = 0, 1, 2 denotes a global regression coefficient for the pth auxiliary variable; β_{pi}, p = 0, 1, 2 denotes a AOI-specific regression coefficient for the pth auxiliary variable, u_{oi}, u_{1i}, and u_{2i} are corresponding AOI-specific random effects, and e_{ij} is a residual error. Superscripts G1, G2, and G3 indicate a model group of AOIs for model fitting (see text for details). The lowest $\hat{\mathrm{AIC}}$ ($\mathrm{c}\hat{\mathrm{AIC}}$) is in bold.

DBH ≥ 36 cm were measured on a 500 m² circular plot, while trees with a DBH < 36 cm and a DBH ≥ 12 cm were measured on a concentric 200 m² circular plot. A total of 5161 plots in the 108 FDs were used for analyses. The number of SNFI field plots in a FD (n_i) varied from 2 to 224 with a mean of 79.

Full census auxiliary data consisted of the Swiss Federal Office of Topography (swisstopo) high precision digital elevation model (swissALTI3D, with a regular grid of 2 m × 2 m). A digital surface model (DSM) was calculated using the maximum value of the airborne laser scanning (ALS) point cloud within each 2 m × 2 m grid cell. The difference in elevation between the surface model and the elevation model was interpreted as the vegetation height. The average point density is reported by swisstopo as 1 point by 2 m² (Artuso et al. 2003). According to swisstopo, the standard error of a vegetation height is 1.5 m.

All pixels classified to forest in swisstopo's Topographic Landscape Model (TLM) and located in one of the 108 FDs were defined as the population under study. At the time of the study, the only available auxiliary variable with full national coverage, a spatial resolution commensurate with the SFNI field plots, and with an expected positive correlation with plot values of stem volume per hectare, was the ALS-based vegetation height (x_{ch}, unit: m). Transforms of x_{ch} (sqrt, log, square, inverse) were explored for their ability to predict y; none qualified as statistically significant (5%) in more than a few (<10) FDs. They are not considered further. The x_{ch} value associated with a SFNI plot is the average vegetation height in 125 2 m × 2 m pixels centered on the 500 m² SFNI plot.

The average of x_{ch} within each FD was computed with GIS software. The forest area in the 108 FDs is approximately 14 000 km².

Vestfold data

Each plot in the Norwegian NFI located in a forest provides a recorded value of stem volume (m³) per ha (y) computed from measurements of DBH and height (HT) and regional volume equations. The NFI design is one of systematic sampling with plot locations on a national grid with a 3 km × 3 km spacing in the part covering Vestfold. Plots are circular with an area of 250 m². There is no stratification in the NFI design covering Vestfold. Again, estimators for simple random sampling are used for reporting on sampling variance.

Forested land in the 14 KOMs in Vestfold defines our study population. The KOMs have been the subject of an earlier SAE study by Breidenbach and Astrup (2012). There are 144 NFI sample plots within the 2184 km² area of Vestfold with data acquired between 2005 and 2010. Vestfold has approximately 123 000 ha of forest dominated by Norway spruce (*Picea abies* L.) and Scots pine (*Pinus sylvestris* L.). Sample size in a KOM (n_i) varied from 1 to 35 with a mean of 20.

Auxiliary data (x) consisting of photogrammetric point cloud metrics of vegetation heights (mean, standard deviation, quartiles, and deciles) in 16 m × 16 m pixels were extracted for all forested land in Vestfold. The point cloud was obtained from overlapping aerial images acquired in 2007 using a Vexcel UltraCamX sensor. A detailed description of the field and remote-sensing data can be found in Breidenbach and Astrup (2012). Preliminary analyses (Breidenbach and Astrup 2012) identified the pixel-level mean vegetation height (x_{ch}) as a statistically significant predictor of biomass ($Mg \cdot ha^{-1}$). We therefore adopted (x_{ch}) as the unique auxiliary variable.

Models and model groups

We consider unit-level linear models for the relationship between y in the jth plot and ith AOI (y_{ij}) as the dependent variable, and indicator variables for an AOI or a group of AOIs and associated auxiliary variable(s) (x_{ij}) as predictors. An indicator variable for the ith AOI takes the value of 1 when a plot resides in the ith AOI or in one of the AOIs that forms a model group for the ith AOI (see below) and zero elsewhere (Särndal et al. 1992, p. 386). Nonlinear and nonparametric models (Opsomer et al. 2008; Pfeffermann 2002; Salvati et al. 2012) can be very effective in SAE applications. Inspection of scatter plots did not, however, suggest any nonlinear relationship between x and y in any of the studied AOIs, neither in the combined SFNI data, nor in the combined Vestfold data.

We settled on a set of 13 linear unit-level models with varying numbers of parameters shared across two or more AOIs. The set is not exhaustive; it reflects reasonable practical alternatives. The models are in Table 1. All models are internal (Mandallaz 2008, section 6.3) in the sense that they are estimated exclusively from the respective NFI volume per ha data (y) and concurrent auxiliary variable(s) of vegetation height (x_{ch}).

All models were, when data was pooled across FDs or across KOMs, statistically significant at the 0.01 level (F-ratio test). Explanatory variables were significant at the 0.05 level in at least one AOI. Models 1 to 10 carry only fixed-effects. Models 11 to 13 include a mixture of fixed and random area effects. Note, models 10 and 13 are similar but for area effects that are fixed in model 10 and random in model 13.

Twelve models (1–9 and 11–13) reflect varying degrees of borrowing strength via parameters shared by more than one AOI (Rao

2003, p. xxi). In model 1, a single model is used for estimation in all AOIs. Model 10 specifies a separate linear model for each AOI. Models 11 to 13 include fixed effects shared by all AOIs and independent random area-level effects (u_i) assumed to be normally distributed (across AOIs) with a mean of zero and a variance σ_u^2. Area-level random effects, as specified here, induce a distance-independent covariance among pairs of y_{ij}s within an AOI. For a random intercept, the covariance is σ_u^2, and for a random slope (u_{pi}) in the pth auxiliary variable (x_{pij}), the covariance between y_{ij} and y_{ik} ($j \neq k$) becomes $x_{pij}x_{pik}\sigma_{up}^2$, where σ_{up}^2 is the among-area variance in the regression coefficient (slope) to x_{pij}.

Models 3 to 9 pursue borrowing strength through formation of model groups (Estevao and Särndal 2004; Lehtonen and Veijanen 2009; Singh and Yuan 2010). A model group is composed of AOIs with presumed similar relationships between y_{ij} and x_{ij}. Members of a model group share a single group-specific model. Three types of groups were formed, indicated by superscripts {G1, G2, or G3}. An AOI can be a member of more than one model group.

A G1 group is formed on the basis of similarities among regression coefficients in AOI-specific models (i.e., model 10). A G1 group for the ith AOI has i and its two nearest regression neighbours as members. Mahalanobis distances (Christensen et al. 2007) between regression coefficients were used to form groups of three AOIs (Baibing 2006). Two FDs (1305 and 2501) in the Swiss data and six municipalities (701, 704, 706, 711, 722, and 725) in the Vestfold data had insufficient sample sizes ($n_i \leq 4$) for this type of grouping. Their G1 groups were based on Euclidean distances between AOI means of y_{ij} and x_{ij}. A total of 79 unique G1 groups were created from the 108 FDs, and eight from the 14 KOMs.

The G2 groups were formed on the basis of spatial proximity. A G2 group for area i has i and its two nearest spatial neighbours as members. Distances were measured between the geographic centers of the AOIs. Nearest neigbours defined by the length of shared borders (Molina et al. 2001) was tried but discarded due to a large number of AOIs with boundaries shared with only one or two AOIs. A total of 93 unique G2 groups were formed from the 108 FDs and 11 from the 14 KOMs.

The Swiss data allowed a third grouping (G3) of FDs based membership in one of five Swiss forest productivity regions (PR). The regions are: PR1 = Jura, PR2 = Plateau, PR3 = Pre-alps, PR4 = Alps, and PR5 = Southern Alps. Although a FDs can straddle a PR boundary, most (>80%) of its forested area is within a single PR. We therefore assigned a FD to the PR holding the largest share of its forest area.

Testing for area effects

Area-level effects and, in particular, random area effects, have an adverse effect on the precision of AOI estimates in SAE applications (Datta et al. 2011). Hence, a priori, it is desirable if these effects can be dispensed without adverse impacts. Recall that if an area effect is zero or nearly zero, a higher level mean can be used with advantages from borrowing strength. To explore the strength of area effects, we conducted a T^2-test of zero area effects in models without area-specific effects (models 1 and 3–8). For a given linear model and m AOIs, the T^2-test is a global lack-of-fit test: $\hat{T}^2 = \sum_{i=1}^{m} \hat{\bar{e}}_i^2 \times \hat{\sigma}_{ei}^{-2}$ where $\hat{\bar{e}}_i^2$ is the square of the difference between the area level mean of y_{ij} ($\bar{y}_i$) and the MA regression estimator of this mean, and $\hat{\sigma}_{ei}^2$ is the OLS estimate of the variance of the regression residuals in the ith AOI. Details and additional motivations for the test are in Datta et al. (2011). For model groups G1 and G2 with a possible overlap in AOI memberships (i.e., models 3 to 6), the T^2-test was modified to the equivalent of a Wald's test (Wald 1941) to accommodate a covariance between area level residuals ($\hat{\bar{e}}_i$) caused by overlapping group memberships. As an example, if area i is a member of its own model group (G_i) and also member of a model group for area j (G_j), the correlation between $\hat{\bar{e}}_i$ and $\hat{\bar{e}}_j$ is, to a first-order approximation, assumed to be $\sqrt{n_i^2 n_{G_i}^{-1} n_{G_j}^{-1}}$ where n_{G_i} and n_{G_j} are the sample sizes in model groups G_i and G_j. The corresponding covariance was computed as, $\hat{\mathrm{cov}}(\hat{\bar{e}}_i, \hat{\bar{e}}_j \mid i \in G_i \cap G_j) \cong n_j^{-1}\hat{\sigma}_i\hat{\sigma}_j \sqrt[-0.5]{n_{G_i}n_{G_j}}$.

Model selection

In pursuit of accurate estimate of the mean stem volume per ha of forested land in the 108 FDs and 14 KOMs, the model presenting the best balance between fit and parsimony was chosen. Akaike's information criterion was used to select the model with the best balance (Akaike and Krishnaiah 1977; Pinheiro and Bates 2000, p. 83). For a linear model without random area effects, Akaike's information criterion is defined as –2 times the log-likelihood of the model residuals plus 2 times the number of parameters in the model. For models with random area effects, we used the conditional AIC (cAIC), which accounts for the estimation of random effects (Vaida and Blanchard 2005). The log-likelihood is interpreted as evidence about a finite population total (mean) under a given model (Royall 2003).

We realize that model selection is enigmatic in the MA paradigm (Lehtonen et al. 2005), where models are only working models (Lehtonen et al. 2003) regarded as fixed in hypothetical replications of the probability sample (Särndal 2007). In most natural resource MA applications, both the model and the explanatory variables are selected by some systematic or ad hoc process. A prudent model selection criterion to guard against optimistic estimates of precision (Chatfield 1995) is therefore in place.

The log-likelihood for Vestfold results is based on the assumption of Gaussian residuals. Anderson–Darling tests of normality of model residuals gave no strong evidence against an assumption of Gaussian residuals in Vestfold (P = 0.31 for pooled residuals, and $P \geq 0.18$ in AOIs with $n_i \geq 6$, and $P \geq 0.47$ in AOIs with $n_i \geq 8$).

In the FD data, we saw 16 rejections of the null hypothesis of normal distributed area-level model residuals. A Student's t-distribution with 10 degrees of freedom was a better choice (no rejection of the null hypothesis with Holm's sequential multiple rejection test at the 5% level (Holm 1979)). Log-likelihood values for the models evaluated for the FDs are therefore based on the assumption of a t_{10} distribution of residuals.

For models 11–13, the likelihood is based on conditional MD estimators (cEBLUPs) (Chambers and Clark 2012, eq. 15.20). We argue for this choice in the section *Estimators of AOI means.* Computational details of cEBLUP are in the Appendix.

For area specific models, the residuals, the log-likelihood and AIC were obtained for each AOI and summed to a population total. For two FDs and six KOMs with $n_i < 4$, the residuals were computed from the best-fit (sense minimum residual variance) group-level (G1, or G2) model for y_{ij}. Residuals from a specific AOI were assumed independent of residuals in all other AOIs. For mixed effects models 11–13, the estimated correlation among y_{ij}s within an AOI was duly considered in the computation of likelihood (Pinheiro and Bates 2000, p. 62).

Inclusion of random effects in a model improves, ceteris-paribus, the log-likelihood of the data. The improvement stems from the implicit covariance among units in an AOI (Pinheiro and Bates 2000, p. 202). However, random effects also flatten the log-likelihood profile (Pawitan et al. 2006), which translates into a larger variance of estimates of fixed parameters (Faes et al. 2009; Fuller 2009, p. 313). Furthermore, the rate of convergence of an estimated variance of a random area effect to its true value depends primarily on the number of AOIs and less on sample size (Datta et al. 2011; Rao 2003, p. 138). These side-effects of adding random effects to a model deserve careful attention.

Since estimators of variance in an estimate of an AOI mean are proportional to the residual variance (Chambers and Clark 2012, (3.5) and (15.23); Särndal et al. 1992, (7.2.11)), it follows that a model selection based on AIC (cAIC) with a penalty levied on the number of model parameters can differ from selections on estimates of accuracy. To illustrate and appreciate the difference, we compare the MA-based estimate of sampling variance in an AOI estimate under the best (sense AIC) linear model to the RMSEs of cEBLUPs (Chambers and Clark 2012, eq. 15.26), and EBLUPs (Rao 2003, eq. 7.2.21) computed with the linear mixed-effects model 13 (the 'full' model).

Model averaging

A large difference in AICs between the models with the second lowest and lowest AIC supports the strategy of selecting a single model as "best" among the set of considered models (Burnham and Anderson 2002, p. 70). However, if more than one model has similar AIC values, a model averaging offers better protection against optimistic estimates of uncertainty (Burnham and Anderson 2002, p. 150). Model averaging begins with an assignment of "Akaike weights" (Burnham and Anderson 2002, p. 75). The weight w_r given to the rth model in a set of R considered models is,

$$(1)\qquad w_r = \frac{\exp(-0.5\Delta_r)}{\sum_{s=1}^{R}\exp(-0.5\Delta_s)},\ \text{where } \Delta_r = \mathrm{AIC}_r - \min_s(\mathrm{AIC}_s)$$

A sum-to-one rescaling of w_r leads to a weight interpretable as "model probability".

The AIC-values in Table 1 are sums taken over, respectively, the Swiss, and the Vestfold AOIs. To compute w_r, $r = 1, \ldots, 13$, we divided the AIC values by the number of AOIs with a sample size of 2 or greater, i.e., 108 FDs and 10 KOMs. The resulting model probabilities are in Table 1. The model weights are subsequently used to generate model-averaged estimators of AOI means and variances (see below).

Small area estimators

Parameters in linear models (1–10) were estimated via OLS techniques (Draper and Smith 1981, ch. 1.2). Empirical best linear unbiased estimators (EBLUE) were used to obtain estimates of the fixed effects in models 11–13, while the variance of random effects were estimated via a restricted maximum likelihood procedure (Pinheiro and Bates 2000, ch. 4).

Model 1 leads to a synthetic MD SAE (Rao 2003, ch. 4.2) in AOIs with insufficient sample sizes for an estimation of an MA variance. For models 2–10, we employed model-assisted (MA) regression estimators of means and variances, as they are design-consistent and nearly design-unbiased (Särndal et al. 1992, p. 273).

For the three linear mixed models, we opted for approximately unbiased conditional EBLUP estimates (cEBLUP) of small area means ($\hat{\bar{y}}_i^{\mathrm{cond}}$) and accuracy (RMSE). Computational details are in the Appendix and in Chambers and Clark (2012, eq. 15.20). Conditioning is on the EBLUP of area-specific random effects. Conditional estimates offer the advantage of: (i) additivity, the area-weighted sum of $\hat{\bar{y}}_i^{\mathrm{cond}}$ is equal to the EBLUP of the overall mean $\hat{\bar{y}}$ and (ii) the ability to write $\hat{\bar{y}}_i^{\mathrm{cond}}$ as a linear combination of the observed values of y in an AOI.

To benchmark small area estimators, we also provide simple random sampling (SRS) estimators for AOI means ($\bar{y}_i$) and variances.

National forest inventories plots in Switzerland and Norway are located on regular national grids. The AOIs form administrative units. They were not explicitly recognized as strata in the inventory designs. In repeated sampling (with a random start of the national grid), the number of plots located in an AOI will be fairly constant. We therefore employed SRS estimators without the extended indicator variable as otherwise recommended when the realized sample size in an AOI follows a Poisson distribution (Lehtonen and Veijanen 2009; Särndal et al. 1992, p. 388).

Hence, the following SRS estimator of the mean for the ith AOI was used (Cochran 1977, p. 20),

$$(2)\qquad \hat{\bar{y}}_i^{\mathrm{srs}} = n_i^{-1}\sum_{j=1}^{n_i} y_{ij}.$$

As a conservative estimator of variance (for systematic sampling) we adopted the variance of a SRS mean,

$$(3)\qquad \hat{\mathrm{var}}\left(\hat{\bar{y}}_i^{\mathrm{srs}}\middle|\, n_i > 1\right) = n_i^{-1}(n_i - 1)^{-1}\sum_{j=1}^{n_i}\left(y_{ij} - \hat{\bar{y}}_i^{\mathrm{srs}}\right)^2.$$

Note that this variance estimator is valid only with $n_i > 1$.

MA estimators (Mandallaz 2013; Särndal et al. 1992, p. 273) of forest district means of y_{ij} ($\bar{y}_i$) were, according to the selected model (10), estimated from,

$$(4)\qquad \hat{\bar{y}}_i^{\mathrm{rgr}} = \hat{\beta}_{0i} + \hat{\beta}_{1i}\bar{x}_i + \hat{\bar{e}}_i$$

where $\bar{x}_i$ is the actual FD mean of the auxiliary variable, $\hat{\bar{e}}_i$ is the average of the regression residuals in the sample from the ith FD (here $\hat{\bar{e}}_i = 0$), and $\hat{\beta}_{0i}$ and $\hat{\beta}_{1i}$ are the OLS estimates of the intercept and slope. Corresponding model-dependent estimators are identical. For the 106 FDs with $n_i \geq 6$, the following MA estimator of variance of $\hat{\bar{y}}_i^{\mathrm{rgr}}$ (Särndal et al. 1992, p. 402) was used,

$$(5)\qquad \hat{\mathrm{var}}_{\mathrm{MA}}\left(\hat{\bar{y}}_i^{\mathrm{rgr}}\right) = n_i^{-1}(n_i - 1)^{-1}\sum_{j=1}^{n_i}\left(\hat{e}_{ij} - \hat{\bar{e}}_i\right)^2.$$

Variances for the two FDs with $n_i = 2$ were estimated with the estimator in (9).

For the 13 KOMs in Vestfold model 1, which has a single common regression model for prediction of y, was chosen. The following least-squares regression was used for estimation in all KOMs,

$$(6)\qquad \hat{y}_{ij} = 26.10 + 0.079x_{ij} + 0.000051x_{ij}^2,\quad \hat{\mathrm{var}}(\hat{y}_{ij}|x_{ij}) = 67.2\ \mathrm{m}^3{\cdot}\mathrm{ha}^{-1},\ \hat{R}^2 = 0.69$$

where $\hat{y}_{ij}$ is the regression estimate of the expected value of y_{ij} ($\mathrm{m}^3{\cdot}\mathrm{ha}^{-1}$) given x_{ij} (m), $\hat{\mathrm{var}}(\hat{y}_{ij}|x_{ij})$ is the conditional variance of $\hat{y}_{ij}$, and $\hat{R}^2$ is the coefficient of determination (viz. the proportion of variance in y_{ij} explained by the model). Model-assisted (and MD) estimators of KOM means of y_{ij} ($\bar{y}_i$) were, in accordance with eq. 6, estimated as,

$$(7)\qquad \hat{\bar{y}}_i^{\mathrm{rgr}} = 26.10 + 0.079\bar{x}_i + 0.000051\overline{(x_i^2)} + \hat{\bar{e}}_i$$

where, as in eq. 4, $\hat{\bar{e}}_i$ is the average of the regression residuals in the sample from the ith KOM. The chosen design-based estimator of variance of $\hat{\bar{y}}_i^{\mathrm{rgr}}$ was (Mandallaz 2013, eq. 21),

$$(8)\qquad \hat{\mathrm{var}}_{\mathrm{MA}}\left(\hat{\bar{y}}_i^{\mathrm{rgr}}\right) = \bar{\mathbf{x}}_i'\hat{\tilde{\boldsymbol{\Sigma}}}_{\hat{\beta}}\bar{\mathbf{x}}_i + n_i^{-1}(n_i - 1)^{-1}\sum_{j=1}^{n_i}\left(\hat{e}_{ij} - \hat{\bar{e}}_i\right)^2,\quad n_i > 1$$

where $\hat{\tilde{\boldsymbol{\Sigma}}}_{\hat{\beta}}$ is an asymptotically consistent estimator of the DB variance of the regression coefficients in eq. 6 (Mandallaz 2013, eq. 15). Fuller (1975) gave an early introduction of this estimator to survey statisticians. For the three KOMs with $n_i = 1$, the last term in eq. 8 was equated to the residual variance in their G2 regression model (model 5 in Table 1).

In the results the error of an estimate is expressed in per cent of the estimated mean, as in, for example,

$$(9)\qquad se_{\mathrm{MA}}\left(\hat{\bar{y}}_i^{\mathrm{rgr}}\right)\% = 100 \times \sqrt{\hat{\mathrm{var}}\left(\hat{\bar{y}}_i^{\mathrm{rgr}}\right)} \times \left(\hat{\bar{y}}_i^{\mathrm{rgr}}\right)^{-1}$$

Details of MD regression estimators (models 1–10) and linear mixed models (11–13) are in Appendix A.

A model-averaged estimator of the mean of y in the ith AOI (Burnham and Anderson 2002, p. 159) was obtained via a bootstrap procedure with B = 10 000 bootstrap replications. Specifically,

$$(10)\qquad \hat{\bar{y}}_i^{\mathrm{avg}} = \frac{1}{B}\sum_{b=1}^{B}\sum_{r=1}^{R}\delta_{r(b)=r}\hat{\bar{y}}_i(r) \simeq \sum_{r=1}^{R} w_r\hat{\bar{y}}_i(r)$$

where $\hat{\bar{y}}_i(r)$ is the target estimate for the ith AOI obtained with the rth model, and $\delta_{r(b)=r}$ is a binary (0,1) indicator variable δ_A taking the value of 1 if event A is true and 0 otherwise. The event $r(b)$ is a random draw of a model from the set of R models with probability proportional to the model probability w_r (see eq. 1). From eq. 10, the following model-averaged estimator of variance is obtained,

$$(11)\qquad \hat{\mathrm{var}}\left(\hat{\bar{y}}_i^{\mathrm{avg}}\right) = \frac{1}{B}\sum_{b=1}^{B}\sum_{r=1}^{R}\delta_{r(b)=r}\Big[\left(\hat{\bar{y}}_i(r) - \hat{\bar{y}}_i^{\mathrm{avg}}\right)^2 \times n_i^{-1} + \hat{\mathrm{var}}\left(\hat{\bar{y}}_i(r)\right)\Big]$$

where $\hat{\mathrm{var}}(\hat{\bar{y}}_i(r))$ is the estimator of uncertainty of the target estimate obtained with model r. With models 1–10, the uncertainty is captured by an estimator of variance and with models 11–13 by a mean squared error.

In eqs. 10 and (11), we intentionally expanded the notation in Burnham and Anderson (2002) to include B bootstrap replications. The expansion highlights the independence of model draws, which, in turn, allows us to ignore otherwise complex among-model covariance structures.

An area-level model with correlated random area effects

Preliminary analyses indicated a statistically significant spatial correlation of random FD effects in the Swiss data (see Appendix B). As a complement to the above unit-level analysis, we sketch how this correlation can be exploited in a Fay–Herriot area-level model and a spatial EBLUP estimator (sEBLUP)(Chandra et al. 2007; Rao 2003, p. 168). Note that this area-level model is not subject to model selection nor to model averaging as it is not nested within the set of unit-level models.

Results

Swiss forest districts

Forest district estimates of $\hat{\bar{y}}_i^{\mathrm{srs}}$ and $\hat{\bar{y}}_i^{\mathrm{rgr}}$ alongside estimates of relative errors are in Table 2 for a random selection of 16 FDs.

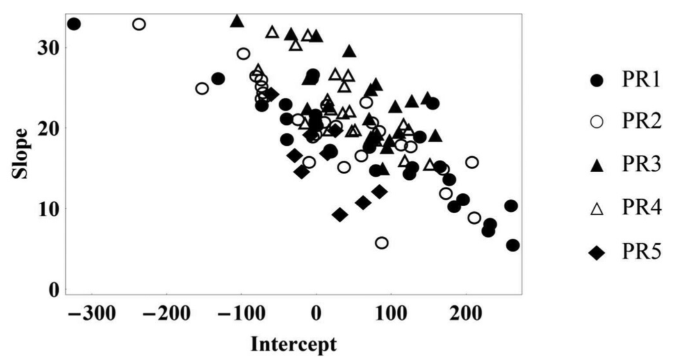
The T^2-tests confirmed the statistical significance at the 5% level or less of the area (FD) effects in models 2–9 (all T^2-statistics were larger than 156, with probabilities less than 0.01 under the null hypothesis of no area effects). A formation of model groups of FDs based on either spatial proximity (G2) or similarity of parameters in FD-specific linear regression models (G1) did not seem to offer important opportunities for borrowing strength by sharing regression parameters among the FDs in a model group. Figure 1 shows a scatter plot of FD-specific slopes and intercepts by PR. There is no apparent clustering, as confirmed by an agglomerative cluster analysis (Everitt et al. 2001), which failed to identify more than a single cluster. Comparable conclusions extend to G1 and G2 groupings.

Table 2. Estimates of mean stem volume ($\bar{y}_i$) ($\mathrm{m^3 \cdot ha^{-1}}$) on forested lands in Swiss forest districts derived from model 10 in Table 1.

District	n_i	$\hat{\bar{y}}_i^{\mathrm{srs}}$	$\hat{\bar{y}}_i^{\mathrm{rgr}}$	$\widehat{se}(\hat{\bar{y}}_i^{\mathrm{srs}})\%$	$\widehat{se}_{\mathrm{MA}}(\hat{\bar{y}}_i^{\mathrm{rgr}})\%$	$\widehat{\mathrm{RMSE}}(\hat{\bar{y}}_i^{\mathrm{cond}})\%$
203	88	395	408	7	6	6
301	44	430	425	9	8	8
403	19	441	386	11	9	10
503	28	453	434	14	11	11
802	9	486	412	15	12	14
1001	21	384	378	14	14	14
1006	32	533	531	11	8	9
1101	36	393	386	10	6	6
1703	64	396	378	9	5	5
1705	60	539	550	8	6	6
1801	94	339	326	6	5	6
1805	206	272	279	4	4	4
1901	77	328	344	8	6	6
2003	40	365	384	10	6	7
2209	22	325	298	15	9	10
2603	51	369	374	8	6	6
Pooled	5161	364	364	1.0	0.7	0.7

Note: table entries are for 16 randomly selected forest districts.
The square root of an estimate of variance in percent of the estimated mean $\hat{y}$ is $\widehat{se}(\hat{y})\%$. The model 13 based estimate of the root mean squared error in percent of $\hat{\bar{y}}_i^{\mathrm{cond}}$ is $\widehat{\mathrm{RMSE}}(\hat{\bar{y}}_i^{\mathrm{cond}})\%$ (see Appendix for details).

Fig. 1. Scatterplot of Swiss forest district-specific intercepts (x-axis) and slopes (y-axis) in model 10 regressions with stem volume per ha (y_{ij}) as the dependent variable.

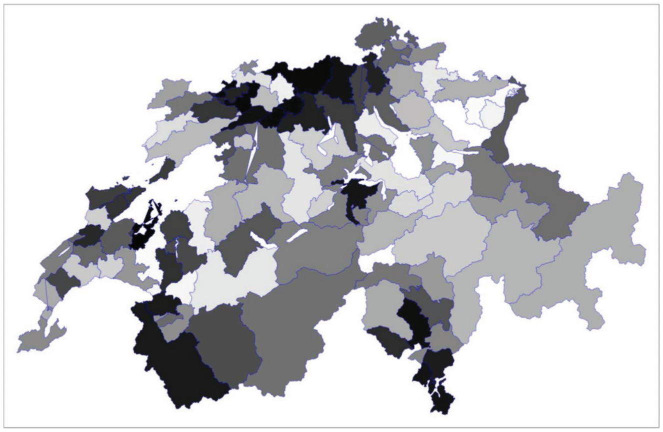


Exploration of spatial trends in model 1 residuals suggested a moderately strong spatial correlation among adjoining FDs. Moran's $\hat{I}$ was 0.13 with a bootstrap 95% confidence interval from 0.09 to 0.17. Moran's $\hat{I}$ for random area effects (model 13) was 0.08 ± 0.03. A map of ranked FD means of model 1 residuals is in Fig. 2. It indicates a mosaic with clusters of high and low values of the average regression residuals. To explore the effects of the suggested spatial correlations, we computed RMSE% (i.e., RMSE in percent of the estimate) of spatial EBLUP estimates of FD means of y obtained with an area-level (Fay–Herriot) model (Molina et al. 2009).

Model 10, i.e., regressions with FD-specific intercepts and slopes, achieved the lowest (overall) AIC (cAIC) value (Table 1). The linear mixed effects model (13) achieved the second lowest AIC value (cAIC). Our MA and MD estimators of a FD mean and variance of stem volume per ha forested land (y) are therefore based on linear regressions with FD-specific intercepts and slopes. The auxiliary variable (x_{ch}) explained between 13% and 72% (mean = 42%) of the within-FD variation in y. For the two FDs with a sample size of two, we estimated for each a G2 regression model (see model 5 in Table 1), i.e., using the FD-specific observations plus observations from the two spatially nearest FDs. The FD mean and variance was computed as for the other FDs.

Model-dependent estimators of variance generated almost identical estimates of relative accuracy $se_{\mathrm{MD}}(\hat{\bar{y}}_i^{\mathrm{rgr}})\%$.

Fig. 2. Relative ranking of 108 Swiss FD means of regression residuals in model 1. The lowest ranked mean is white and the largest is black.



Estimates $\hat{\bar{y}}_i^{\text{srs}}$ of the mean volume per ha on forested land in a FD varied from 160 to 653 m³·ha⁻¹, while $\hat{\bar{y}}_i^{\text{rgr}}$ varied from 163 to 607 m³·ha⁻¹. The two sets of estimates were strongly correlated (0.93), yet Kendall's rank correlation was no more than 0.75, and the average rank change between the two sets of estimates was 9. In the combined sample of all FDs, the average of x_{ch} was 15.3 m, or 1.0 m above the census mean of 14.3 m. Despite the difference in the means of the auxiliary variable, the FD area-weighted estimates for the study population based on $\hat{\bar{y}}_i^{\text{rgr}}$ and $\hat{\bar{y}}_i^{\text{srs}}$ were close (363 and 364 m³·ha⁻¹, respectively).

Exploiting the relationship between the auxiliary variable (x_{ch}) and the target variable (y) reduced the average relative error of a FD estimate of the mean stem volume per ha by 23% (range 0% to 52%). Estimates of the relative errors of $\hat{\bar{y}}_i^{\text{srs}}$ varied from 4.2% to 23.7%, whereas the estimated relative errors of $\hat{\bar{y}}_i^{\text{rgr}}$ were between 3.5% and 24.5%. In five FDs (1001, 1301, 2204, 2206, and 2016), the regression was not statistically significant, and the relative error was almost at par with the results from SRS. With MA estimators of variance, the number of FDs with an estimated relative error less than 10% was 87. The corresponding count for the SRS estimates was 55.

Had we selected the linear mixed effects model (13) and cEBLUP estimators of FD means ($\hat{\bar{y}}_i^{\text{cond}}$), our average estimate of uncertainty would have been very similar. In 91 FDs, the MD conditional approach to SAE would result in a larger estimate of uncertainty (see selected examples in Table 2). For orientation, simple averages of the two sets of estimates of relative uncertainty were, for the 106 FDs with $n_i \geq 6$, 8.3% for the MA estimates and 9.7% for the cEBLUP estimates.

We have argued for a model-selection based on AIC (cAIC) and cEBLUP estimators for mixed effects models. However, Fig. 3 illustrates that the relative accuracy of an EBLUP of an FD mean can be much higher than the estimated accuracy of $\hat{\bar{y}}_i^{\text{rgr}}$ and $\hat{\bar{y}}_i^{\text{cond}}$. When averaged over FDs with $n_i \geq 6$, the RMSE% of EBLUPs was 5.7%, and the average for cEBLUPs was 8.0%.

Model averaging is an option to consider in the Swiss SAE. Model 10 selected by AIC is only given a relative Akaike weight of 26% (see Table 2), while the two runner up models (5 and 13) have relative weights of 14%. Model averaging had an effect on the 106 $n_i > 5$ AOI estimates of mean and variance. The average absolute change in an AOI mean was 5% with the largest changes concentrated in AOIs with the smallest sample sizes. Model-averaged SAEs of relative errors were, on average, 8% larger than with model 10 (Fig. 3). Model 10 and model averaged SAEs were strongly correlated (>0.95).

Vestfold municipalities

Estimates of $\hat{\bar{y}}_i^{\text{srs}}$ and $\hat{\bar{y}}_i^{\text{rgr}}$ are in Table 3 with estimates of relative errors.

The lowest AIC (cAIC) was obtained for a single global second-degree polynomial in x_{ch} (model 1 in the Vestfold column of Table 1). The T^2-test statistic under the hypothesis of no area effects in this model was $\hat{T}^2_{14-3} = 18.3$, $P(\hat{T}^2_{14-3} \mid H_0) = 0.08$. Tests with model groups G1 and G2 levels corroborated the absence of KOM-specific effects ($\hat{T}^2 < 7.2$, $P > 0.14$). The random-effects model (13) was the runner-up in cAIC, but all variance components of the random effects were very small, as exemplified by the ratio of the among- to within-KOM variances (Fuller 2009, p. 312) of per hectare stem volume of just 0.0027. Also, Moran's I was negative $\hat{I} = -0.22 \pm 0.08$ for model 1 residuals. A map (not shown) of predicted random area effects equally failed to convey any positive correlation among pairs of KOMs sharing a border. Attempts to compute sEBLUPs failed due to singularities of matrices to be inverted.

Estimates of the mean stem volume per ha in forest land of a KOM obtained directly from the sample ($\hat{\bar{y}}_i^{\text{srs}}$) varied from a low of 32 m³·ha⁻¹ to a high of 225 m³·ha⁻¹. A much narrower range (85–169) was, as expected, seen in $\hat{\bar{y}}_i^{\text{rgr}}$. With increasing sample size, the difference between $\hat{\bar{y}}_i^{\text{srs}}$ and $\hat{\bar{y}}_i^{\text{rgr}}$ shrank. For small KOM sample sizes, the differences were typically quite large (20% to 26%). Overall, the area-weighted average of $\hat{\bar{y}}_i^{\text{srs}}$ was higher than the area weighted average of $\hat{\bar{y}}_i^{\text{rgr}}$. The difference (10 m³·ha⁻¹) is attributed to a difference of 0.8 m between the mean canopy height in the 144 NFI plots and the area-weighted mean of the known KOM means of x_{ch}. A relative precision of $\hat{\bar{y}}_i^{\text{srs}}$ below 15% was only achieved in KOM 709 with a sample size of 35. With $\hat{\bar{y}}_i^{\text{rgr}}$, six KOMs qualified in this regard (Table 3). A sample size below four in a KOM gave estimates of $\bar{y}_i$ with relative errors in excess of 30%. The use of x_{ch} and its square as auxiliary variables lowered the standard errors of an estimated KOM mean of volume per ha in forested land by approximately 22%.

An MD approach to inference with model 1 generated, as expected, almost identical estimates of relative accuracy $se_{\text{MD}}(\hat{\bar{y}}_i^{\text{rgr}})\%$. Selecting the linear mixed effects model (13), with cEBLUP estimators of KOM means, would have generated less accurate estimates (Table 3 and Fig. 4). The mean of RMSE$(\hat{\bar{y}}_i^{\text{cond}})$% for KOMs with sample sizes of 2 or greater was 39%.

Estimates of accuracy would have been improved (average 16%) had we adopted EBLUP estimators with model 13 (Fig. 4).

Model averaging is also an option to consider in Vestfold. Model (1) selected by AIC is only given a relative Akaike weight of 29% (see Table 2), while the two runner up models (2 and 13) have relative weights of 17% and 15%, respectively. Model averaging had an effect on the 8 $n_i > 5$ AOI estimates of mean and variance. The average absolute change in an AOI mean was 10%, with the largest changes concentrated in AOIs with small sample sizes. Model-averaged SAEs of relative errors were, on average, 41% larger than with model 1 (Fig. 4). Model 1 and model averaged SAEs were strongly correlated (>0.92).

Discussion

National forest inventories provide accurate national and regional estimates of important forest resource variables (Bechtold and Patterson 2005; Gjertsen et al. 1999). National inventories are typically expensive and demand the best possible use of collected data. Small area estimation is an obvious extension, especially with the increased availability of relatively cheap auxiliary variables known for the entire AOI or at least for a large proportion of the AOI (McRoberts and Tomppo 2007; Tomppo 1991). ALS data have been particularly useful in SAE applications, as they provide proxy estimates of canopy height, a variable known to correlate well, at the plot-level, with mean tree heights, per hectare stem volume, per hectare basal area, and per hectare aboveground bio-

Fig. 3. Relative accuracy (RMSE%) of estimates of FD means of stem volume per ha on forested land plotted against area-level sample sizes (n_i). Model 13 was used for cEBLUPs and EBLUPs, while sEBLUPs were obtained from an area level Fay–Herriot model with spatial correlation of area effects. For MA estimates derived from model 10, the relative accuracy is equal to the relative sampling error. The dashed line is the nonlinear least squares trend in (RMSE%) of model averaged FD means.

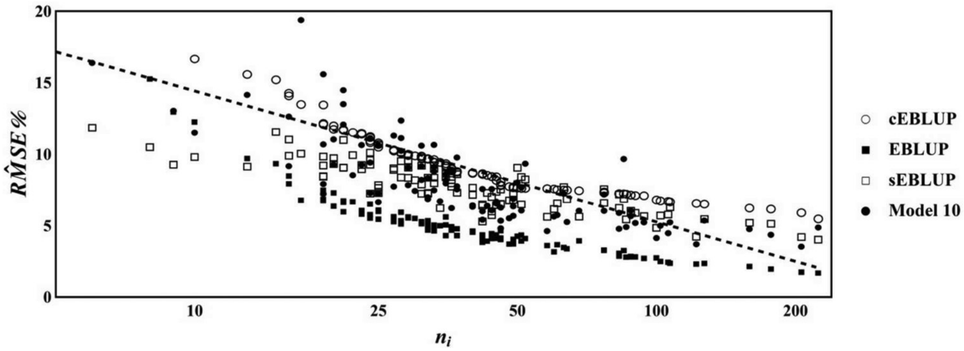


Table 3. Estimators of mean stem volume per ha ($\bar{y}_i$) on forested land in 14 Vestfold municipalities (KOMs).

KOM	n_i	$\hat{\bar{y}}_i^{srs}$	$\hat{\bar{y}}_i^{rgr}$	$\widehat{se}(\hat{\bar{y}}_i^{srs})\%$	$\widehat{se}_{MA}(\hat{\bar{y}}_i^{rgr})\%$	$\widehat{RMSE}(\hat{\bar{y}}_i^{cond})\%$
701	1	94	118	n.a.	77	n.a.
702	6	126	94	43	36	42
704	3	225	132	24	38	57
706	2	64	122	64	57	88
709	35	142	131	14	9	18
711	4	109	159	25	26	56
713	17	193	160	26	12	22
714	12	122	112	17	18	29
716	12	129	132	19	15	34
719	13	166	120	18	15	26
720	8	186	169	26	15	36
722	1	32	85	n.a.	100	n.a.
723	1	158	154	n.a.	54	n.a.
728	29	114	119	15	11	21
Pooled	144	139	129	7	5	6

Note: The square root of an estimate of variance in percent of the estimated mean $\hat{\bar{y}}$ is $\widehat{se}(\hat{\bar{y}})\%$. The model 13 based estimate of the root mean squared error in percent of $\hat{\bar{y}}_i^{cond}$ is $\widehat{RMSE}(\hat{\bar{y}}_i^{cond})\%$ (see Appendix for details).

mass of trees (Magnussen et al. 2013; McRoberts et al. 2010; Wulder et al. 2012).

In our case, a single auxiliary variable of mean canopy height reduced the relative uncertainty in an MA AOI estimate of volume per ha in forested land by about 20%, an appreciable result, although the reported levels of errors may not, in isolation, support effective decision making for an AOI.

In forest survey data, plot-level y-variables like mean tree height, age, species composition, site-index, and stem density are important explanatory variables for the stem volume per ha in a forest inventory plot (Vanclay and Skovsgaard 1997). AOI-level estimates of these variables, or at least remotely sensed proxies of these variables, would be needed to improve the precision of AOI estimates of the mean volume per ha forest land.

Additional improvements in estimates of accuracy could have been realized with an area-by-area model selection approach. However, the issue of model selection bias (Chatfield 1995; Draper 1995) quickly becomes an important issue if the approach is extended beyond a single AOI. Mixed-effects models would seem most attractive when inference is focused on a single AOI with a small sample size (Ghosh and Rao 1994).

At first glance, a 20% reduction in the relative MA standard error may seem a modest achievement until one realizes that a parallel reduction in a SRS estimate of error would require a 60% increase in sample size (Cochran 1977, p. 77).

It is clear that with our model selection criterion (AIC, cAIC) and our choice of the cEBLUP estimator for linear mixed effects models, we have forgone potentially significant improvements in estimates of accuracy. As shown, EBLUP estimates of accuracy can be particularly attractive (Breidenbach and Astrup 2012; Goerndt et al. 2011, 2013; Katila 2006; McRoberts 2011), in particular in SAE studies with evidence of a strong spatial auto-correlation process (Chandra et al. 2012; Petrucci et al. 2005; Pfeffermann 2013; Singh et al. 2005; Sperlich and José Lombardía 2010). Geographically weighted regression estimators also appear attractive in terms of estimated RMSEs (Petrucci et al. 2005; Salvati et al. 2012).

Further improvements in our MA-based estimates of accuracy could have been achieved with optimal calibration estimators (Lehtonen and Vejanen 2012; Särndal 2007; Wu 2003), but several issues of when and how to apply these estimators requires careful considerations.

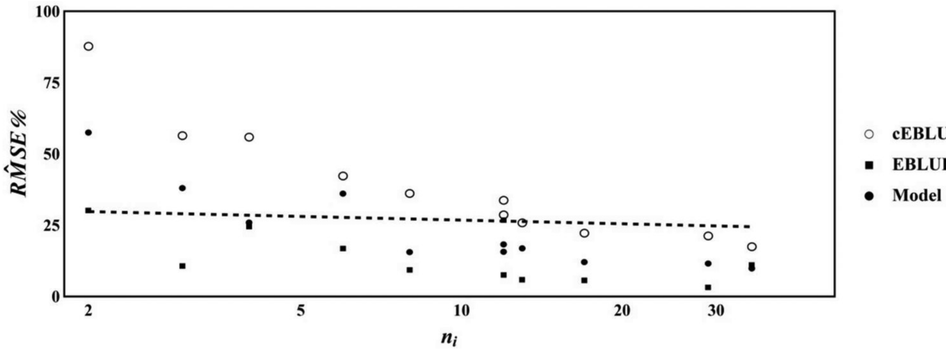
Selecting a model and an estimator based on a potentially highly imprecise estimate of a RMSE (Longford 2005, p. 252) can lead to over-fitting and a false sense of optimism (Hastie et al. 2005, ch. 7.4). Pooling of data for the EBLUP estimation of the within-area variance is likely the main reason behind seemingly favorable EBLUP estimates of accuracy (Jiang and Lahiri 2006*b*; Rao 2003, p. 109).

It is important to use the cAIC criterion for models with random area effects because the log-likelihood of the data can be lowered considerably by simply adding a few random effects (Mayr et al. 2012; Vaida and Blanchard 2005). Had we used the unconstrained AIC, we would have selected the mixed effects model 13 for both the FDs and the KOMs.

An outstanding question is whether AIC (cAIC) provides sufficient protection against optimistic estimates of variance. When models are not fixed a priori, they become random, causing extra variance in estimates across replicated sampling. Model averaging is intended to capture this variance. In our SAE problems, no single model achieved a relative Akaike weight (Burnham and Anderson 2002, p. 75) greater than 0.3, and the second and third highest weights were greater than 10%. In this situation, model averaging may be an option to consider. In the Swiss case, a model averaging increased the average estimate of error by a modest 8%, a reflection on the generally large sample sizes and the relatively low errors of a SAE. In contrast, much smaller sample sizes in Vestfold set the stage for a larger impact of model averaging. To wit, both changes in means and increases in estimates of error were considerably greater in Vestfold. Large discrepancies between estimates from a single model and from a model averaging suggest an improved SAE inference following a model averaging.

In defense of cEBLUP estimators for mixed effects models, we echo Chambers and Clark (2012) when they say "an unconditional approach to SAE takes no account of the actual value of the area effect in any particular area, and how this could affect the uncertainty about the true value of the area mean. [So] the unconditional approach is driven essentially by the variability of area

Fig. 4. Relative accuracy (RMSE%) of estimates of KOM means of stem volume per ha on forested land plotted against area-level sample sizes (n_i). Model 13 was used for cEBLUPs. For estimates derived from model 1, the relative accuracy is equal to the relative sampling error. The dashed line is the nonlinear least squares trend in (RMSE%) of model averaged KOM means.



effects across all areas. It can be argued that this is not necessarily the best way to measure the uncertainty in the estimate for any particular area, since this uncertainty is then being measured relative to situations where the effect for the area of interest is completely different. … What is required is a measure of uncertainty conditional on the actual values of the area effects in the different small areas, rather than an unconditional measure that averages over these effects." Conditional EBLUP estimates $\hat{\bar{y}}_i^{\text{cond}}$ tracks observed sample means $\hat{\bar{y}}_i^{\text{srs}}$ much closer than unconditional EBLUPs (e.g., Rao 2003, p. 96) and $\hat{\text{var}}(\hat{\bar{y}}_i^{\text{cond}})$ will be both larger and much more variable than an unconditional estimate of variance (e.g., Rao 2003, p. 135). This was confirmed with both the Swiss and the Vestfold data. Conditional estimates of relative errors were, for AOIs with a sample size of four or greater, approximately 1.3 times larger (Swiss FDs) and 1.8 times larger (Vestfold KOMs) than corresponding unconditional counterparts. The ratio of the largest to the smallest estimate of a conditional error was approximately three times larger than the corresponding ratio of unconditional errors.

In forestry, the choice of cEBLUP can also be argued on the grounds that within-area variances reflect forest conditions and forest structures. An assumption in the MSE of an EBLUP estimate is that the within-area variance is proportional to the inverse of the area sample size, is therefore often not satisfied. A third argument is that cEBLUP SAE estimates are calibrated to the EBLUP estimator for the combined set of AOIs. This benchmarking should be a requirement for all SAEs derived from a probability sample (Jiang and Lahiri 2006*a*, 2006*b*; Jiang et al. 2011). Fuller (2009) details a "compromise" cEBLUP benchmarked on the auxiliary variable.

For agencies with a preference for a DB inference (Hansen et al. 1953, p. 35; Little 2004), the estimators by Mandallaz (Mandallaz 2013; Mandallaz et al. 2013) become attractive for SAE problems. When several model alternatives present themselves, we argue for the model selection procedure used here, well knowing that it remains controversial in the DB paradigm.

Acknowledgements

We are grateful to the Swiss Federal Institute for Forest, Snow, and Landscape Research WSL for their support of this study and to the Norwegian Forest and Landscape Institute for their cooperation and data sharing. Three journal reviewers and the Associate Editor provided constructive criticism and helpful suggestions to earlier versions of our manuscript.

References

Akaike, H., and Krishnaiah, P.R. 1977. On entropy minimization principle. *In* Applications of statistics. North Holland, Amsterdam. pp. 27–42.

Artuso, R., Bovet, S., and Streilein, A. 2003. Practical Methods for the Verification of Countrywide Terrain and Surface Models. *In* 3-D Reconstruction from Airborne Laserscanner and InSAR Data. ISPRS Working Group III/3, Dredsen, GER.

Baibing, L. 2006. A new approach to cluster analysis: the clustering-function-based method. J. R. Stat. Soc. Series B Stat. Methodol. **68**(3): 457–476. doi:10.1111/j.1467-9868.2006.00549.x.

Bechtold, W.A., and Patterson, P.L. 2005. The enhanced forest inventory and analysis program - National sampling design and estimation procedures. Gen. Tech. Rep. SRS-80.

Brändli, U.-B. 2010. Schweizerisches Landesforstinventar. Ergebnisse der dritten Erhebung 2004–2006.

Breidenbach, J., and Astrup, R. 2012. Small area estimation of forest attributes in the Norwegian National Forest Inventory. Eur. J. For. Res. **131**(4): 1255–1267. doi:10.1007/s10342-012-0596-7.

Breidt, F.J. 2004. Small area estimation for natural resource surveys. *In* Monitoring Science & Technology Symposium, Denver, CO.

Burnham, K.P., and Anderson, D.R. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2nd ed. Springer, New York.

Chambers, R.L., and Clark, R.G. 2012. An introduction to model-based survey sampling with applications. Oxford University Press, New York.

Chandra, H., Salvati, N., and Chambers, R. 2007. Small Area Estimation for Spatially Correlated Populations – A Comparison of Direct and Indirect Model-Based Methods. Working Paper.

Chandra, H., Salvati, N., Chambers, R., and Tzavidis, N. 2012. Small area estimation under spatial nonstationarity. Comput. Stat. Data Anal. **56**(10): 2875–2888. doi:10.1016/j.csda.2012.02.006.

Chatfield, C. 1995. Model Uncertainty, Data Mining and Statistical Inference. J. R. Stat. Soc. Series A Stat. Soc. **158**(3): 419–466. doi:10.2307/2983440.

Christensen, W.F., Dillner, A.M., Schauer, J.J., and Reese, C.S. 2007. Clustering composition vectors using uncertainty information. Environmetrics, **18**(8): 859–869. doi:10.1002/env.847.

Claeskens, G., and Hjort, N.L. 2008. Model selection and model averaging. Cambridge University Press, Cambridge.

Cliff, A.D., and Ord, J.K. 1981. Spatial Processes. Pion, London.

Cochran, W.G. 1977. Sampling techniques. Wiley, New York.

Corona, P., Chirici, G., and Marchetti, M. 2002. Forest ecosystem inventory and monitoring as a framework for terrestrial natural renewable resource survey programmes. Plant Biosyst. **136**(1): 69–82. doi:10.1080/11263500212331358531.

Cressie, N.A.C. 1993. Statistics for spatial data. Revised edition. 2nd ed. Wiley, New York.

Datta, D.S., and Lahiri, P. 2000. A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. Stat. Sin. **10**(2): 613–627.

Datta, G.S. 2009. Chapter 32 – Model-Based Approach to Small Area Estimation. *In* Handbook of Statistics. *Edited by* C.R. Rao. Elsevier. pp. 251–288.

Datta, G.S., Hall, P., and Mandal, A. 2011. Model Selection by Testing for the Presence of Small-Area Effects, and Application to Area-Level Data. J. Am. Stat. Assoc. **106**(493): 362–374. doi:10.1198/jasa.2011.tm10036.

Deville, J.-C. 1999. Variance estimation for complex statistics and estimators: Linearization and residual techniques. Surv. Meth. **25**(2): 193–203.

Draper, D. 1995. Assessment and propagation of model uncertainty. J. R. Stat. Soc. Series B Stat. Methodol. **57**(1): 45–97.

Draper, N.R., and Smith, H. 1981. Applied Regression Analysis. Wiley, New York.

Estevao, V.M., and Särndal, C.-E. 2004. Borrowing strength is not the best technique within a wide class of design-consistent domain estimators. J. Off. Stat. **20**(4): 645–669.

Everitt, B.S., Landau, S., and Leese, M. 2001. Cluster Analysis. 4th ed. Arnold, London.

Faes, C., Molenberghs, G., Aerts, M., Verbeke, G., and Kenward, M.G. 2009. The

Effective Sample Size and an Alternative Small-Sample Degrees-of-Freedom Method. Am. Stat. **63**(4): 389–399. doi:10.1198/tast.2009.08196.

Freeman, E., and Moisen, G. 2007. Evaluating Kriging as a Tool to Improve Moderate Resolution Maps of Forest Biomass. Environ. Monit. Assess. **128**(1): 395–410. doi:10.1007/s10661-006-9322-6.

Fuller, W.A. 1975. Regression analysis for sample survey. Sankhya, **37**(3): 117–132.

Fuller, W.A. 2009. Sampling Statistics. Wiley, New York.

Gallaun, H., Zanchi, G., Nabuurs, G.J., Hengeveld, G., Schardt, M., and Verkerk, P.J. 2010. EU-wide maps of growing stock and above-ground biomass in forests based on remote sensing and field measurements. For. Ecol. Manage. **260**(3): 252–261. doi:10.1016/j.foreco.2009.10.011.

Ghosh, M., and Rao, J.N.K. 1994. Small area estimation: an appraisal. Stat. Sci. **9**: 55–93. doi:10.1214/ss/1177010647.

Gillespie, A.J. 1999. Rationale for a national annual forest inventory program. J. For. **97**(12): 16–20.

Gjertsen, A.K., Tomppo, E., and Tomter, S. 1999. National forest inventory in Norway: Using sample plots, digital maps, and satellite images. *In*. IEEE.

Goerndt, M.E., Monleon, V.J., and Temesgen, H. 2011. A comparison of small-area estimation techniques to estimate selected stand attributes using LiDAR-derived auxiliary variables. Can. J. For. Res. **41**(6): 1189–1201. doi:10.1139/x11-033.

Goerndt, M.E., Monleon, V.J., and Temesgen, H. 2013. Small-Area Estimation of County-Level Forest Attributes Using Ground Data and Remote Sensed Auxiliary Information. For. Sci. **59**(5): 536–548. doi:10.5849/forsci.12-073.

Gregoire, T.G. 1998. Design-based and model-based inference in survey sampling: Appreciating the difference. Can. J. For. Res. **28**(10): 1429–1447. doi:10.1139/x98-166.

Hansen, M.H., Hurwitz, W.N., and Madow, W.G. 1953. Sample survey methods and theory. John Wiley, New York.

Hastie, T., Tibshirani, R., Friedman, J., and Franklin, J. 2005. The elements of statistical learning: data mining, inference and prediction. Math. Intel. **27**(2): 83–85. doi:10.1007/BF02985802.

Holm, S. 1979. A simple sequentially rejective multiple test procedure. Scand. J. Stat. **6**: 65–70.

Jiang, J., and Lahiri, P. 2006*a*. Estimation of finite population domain means: A model-assisted empirical best prediction approach. J. Am. Stat. Assoc. **101**(473): 301–311. doi:10.1198/016214505000000790.

Jiang, J., and Lahiri, P. 2006*b*. Mixed Model Prediction and Small Area Estimation. Test, **15**(1): 1–96. doi:10.1007/BF02595419.

Jiang, J., Nguyen, T., and Rao, J.S. 2011. Best Predictive Small Area Estimation. J. Am. Stat. Assoc. **106**(494): 732–745. doi:10.1198/jasa.2011.tm10221.

Kangas, A. 1996. Small-area estimates using model-based methods. Can. J. For. Res. **26**(5): 758–766. doi:10.1139/x26-085.

Katila, M. 2006. Empirical errors of small area estimates from the multisource national forest inventory in eastern Finland. Silv. Fenn. **40**(4): 729–742.

Katila, M., and Tomppo, E. 2006. Sampling, simulation on multi-source output forest maps – an application for small areas. *Edited by* M.E. Caetano and M. Painho. Portugese Geographic Institute, Lisbon.

Kaufmann, E. 1999. Vorrat, Zuwachs, Nutzung. *In* Schweiyerisches Landesforstinventar – Methoden und Modelle der Zweitaufnahme 1993–1995. *Edited by* P. Brassel and H. Lischke. Eidgenössissche Forschungsanstalt Wald Schnee Landschaft, Birmensdorf. pp. 162–196.

Lehtonen, R., and Veijanen, A. 2009. Design-based Methods of Estimation for Domains and Small Areas. *In* Handbook of Statistics. *Edited by* C.R. Rao. Elsevier. pp. 219–249.

Lehtonen, R., and Vejanen, A. 2012. Small area poverty estimation by model calibration. J. Indian Soc. Agric. Stat. **66**(1): 125–133.

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

Lehtonen, R., Särndal, C.-E., and Veijanen, A. 2005. Does the model Matter? Comparing model-Assisted and Model-Dependent estimators of class frequencies for domains. Statistics in Transition, **7**(3): 649–673.

Little, R.J.A. 2004. To model or not to model? Competing modes of inference for finite population sampling. J. Am. Stat. Assoc. **99**(466): 546–556. doi:10.1198/016214504000000467.

Lohr, S.L., and Prasad, N.G.N. 2003. Small area estimation with auxiliary survey data. Can. J. Stat. **31**(4): 383–396. doi:10.2307/3315852.

Longford, N.T. 2005. Missing data and small-area estimation. Springer, New York.

Magnussen, S., Naesset, E., and Gobakken, T. 2013. Prediction of tree-size distributions and inventory variables from cumulants of canopy height distributions. Forestry, **86**(5): 583–595. doi:10.1093/forestry/cpt022.

Maltamo, M., Packalén, P., Suvanto, A., Korhonen, K.T., Mehtätalo, L., and Hyvönen, P. 2009. Combining ALS and NFI training data for forest management planning: A case study in Kuortane, Western Finland. Eur. J. For. Res. **128**(3): 305–317. doi:10.1007/s10342-009-0266-6.

Mandallaz, D. 2008. Sampling techniques for forest inventories. Chapman and Hall, Boca Raton, Florida.

Mandallaz, D. 2013. Design-based properties of some small-area estimators in forest inventory with two-phase sampling. Can. J. For. Res. **43**(5): 441–449. doi:10.1139/cjfr-2012-0381.

Mandallaz, D., Breschan, J., and Hill, A. 2013. New regression estimators in forest inventories with two-phase sampling and partially exhaustive information: a design-based Monte Carlo approach with applications to small-area estimation. Can. J. For. Res. **43**(11): 1023–1031. doi:10.1139/cjfr-2013-0181.

Mayr, A., Fenske, N., Hofner, B., Kneib, T., and Schmid, M. 2012. Generalized additive models for location, scale and shape for high dimensional data-a flexible approach based on boosting. J. R. Stat. Soc. Series C Appl. Stat. **61**(3): 403–427. doi:10.1111/j.1467-9876.2011.01033.x.

McRoberts, R.E. 2011. 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., and Tomppo, E.O. 2007. Remote sensing support for national forest inventories. Remote Sens. Environ. **110**: 412–419. doi:10.1016/j.rse.2006.09.034.

McRoberts, R.E., Tomppo, E.O., and Næsset, E. 2010. Advances and emerging issues in national forest inventories. Scand. J. For. Res. **25**(4): 368–381. doi:10.1080/02827581.2010.496739.

Molina, E.A., Smith, T.M.F., and Sugden, R.A. 2001. Modelling Overdispersion for Complex Survey Data. Int. Stat. Rev. **69**(3): 373–384. doi:10.1111/j.1751-5823.2001.tb00464.x.

Molina, I., Salvati, N., and Pratesi, M. 2009. Bootstrap for estimating the MSE of the spatial EBLUP. Comput. Stat. **24**(3): 441–458. doi:10.1007/s00180-008-0138-4.

Olsen, A.R., Sedransk, J., Edwards, D., Gotway, C.A., Liggett, W., Rathburn, S., Reckhow, K.H., and Young, L.J. 1999. Statistical issues for monitoring ecological and natural resources in the United States. Environ. Monit. Assess. **54**: 1–45. doi:10.1023/A:1005823911258.

Opsomer, J.D., Claeskens, G., Ranalli, M.G., Kauermann, G., and Breidt, F.J. 2008. Non-parametric small area estimation using penalized spline regression. J. R. Stat. Soc. Series B Stat. Methodol. **70**(1): 265–286. doi:10.1111/j.1467-9868.2007.00635.x.

Pawitan, Y., Calza, S., and Ploner, A. 2006. Estimation of false discovery proportion under general dependence. Bioinformatics, **22**(24): 3025–3031. doi:10.1093/bioinformatics/btl527.

Petrucci, A., Pratesi, M., and Salvati, N. 2005. Geographic information in small area estimation: Small area models and spatially correlated random area effects. Statistics in Transition, **7**(3): 609–623.

Pfeffermann, D. 2002. Small area estimation - New developments and directions. Int. Stat. Rev. **70**(1): 125–143. doi:10.1111/j.1751-5823.2002.tb00352.x.

Pfeffermann, D. 2013. New important developments in small area estimation. Stat. Sci. **28**: 40–68. doi:10.1214/12-STS395.

Pinheiro, J.C., and Bates, D.M. 2000. Mixed-effects models in S and S-plus. Springer, New York.

Pratesi, M., and Salvati, N. 2008. Small area estimation: the EBLUP estimator based on spatially correlated random area effects. Stat. Methods Appl. **17**(1): 113–141. doi:10.1007/s10260-007-0061-9.

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

Royall, R. 2003. Interpreting a sample as evidence about a finite population. *In* Analysis of survey data. *Edited by* R.L. Chambers and C.J. Skinner. Wiley, Chichester. pp. 59–72.

Saei, A., and Chambers, R. 2003. Small area estimation: A review of methods-based on the application of mixed models. Southampton Statistical Sciences Research Institute.

Salvati, N., Tzavidis, N., Pratesi, M., and Chambers, R. 2012. Small area estimation via M-quantile geographically weighted regression. Test, **21**(1): 1–28. doi:10.1007/s11749-010-0231-1.

Särndal, C.-E. 2007. The calibration approach in survey theory and practice. Surv. Meth. **33**(2): 99–119.

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

Shen, X.T., Huang, H.C., and Ye, J. 2004. Inference after model selection. J. Am. Stat. Assoc. **99**(467): 751–762. doi:10.1198/016214504000001097.

Singh, A.C., and Yuan, P. 2010. Building-block BLUPs for aggregate level small area estimation for survey data. *In* Joint Statistical Meeting. American Statistical Association. p. 14.

Singh, B.B., Shukla, G.K., and Kundu, D. 2005. Spatio-temporal models in small area estimation. Surv. Meth. **31**(2): 183–195.

Sperlich, S., and José Lombardía, M. 2010. Local polynomial inference for small area statistics: estimation, validation and prediction. J. Nonparametr. Stat. **22**(5): 633–648. doi:10.1080/10485250903311607.

Ståhl, G., Allard, A., Esseen, P.A., Glimskär, A., Ringvall, A., Svensson, J., Sundquist, S., Christensen, P., Torell, Å.G., Högström, M., Lagerqvist, K., Marklund, L., Nilsson, B., and Inghe, O. 2011. National Inventory of Landscapes in Sweden (NILS)-scope, design, and experiences from establishing a multiscale biodiversity monitoring system. Env. Monit. Assess. **173**(1–4): 579–595. doi:10.1007/s10661-010-1406-7.

Steinmann, K., Mandallaz, D., Ginzler, C., and Lanz, A. 2013. Small area estimations of proportion of forest and timber volume combining Lidar data and stereo aerial images with terrestrial data. Scand. J. For. Res. **28**(4): 373–385. doi:10.1080/02827581.2012.754936.

Tomppo, E. 1991. Satellite image-based national forest inventory of Finland. *In* Proceedings of the symposium on global and environmental monitoring, techniques and impacts. ISPRS, Victoria BC. pp. 419–424.

Tomppo, E., Olsson, H., Stahl, G., Nilsson, M., Hagner, O., and Katila, M. 2008. Combining national forest inventory field plots and remote sensing data for

forest databases. Remote Sens. Environ. **112**(5): 1982–1999. doi:10.1016/j.rse.2007.03.032.
Tomppo, E., Gschwantner, T., Lawrence, M., McRoberts, R., Gabler, K., Schadauer, K., Vidal, C., Lanz, A., Ståhl, G., and Cienciala, E. 2010. National Forest Inventories. *In* Pathways for Common Reporting. European Science Foundation.
Vaida, F., and Blanchard, S. 2005. Conditional Akaike information for mixed-effects models. Biometrika, **92**(2): 351–370. doi:10.1093/biomet/92.2.351.
Vanclay, J.K., and Skovsgaard, J.P. 1997. Evaluating forest growth models. Ecol. Model. **98**(1): 1–12. doi:10.1016/S0304-3800(96)01932-1.
Wald, A. 1941. Asymptotically most powerful test of statistical hypotheses. Ann. Math. Stat. **12**: 1–19. doi:10.1214/aoms/1177731783.
Wolfram, S. 1999. The Mathematica Book. Wolfram Media/Cambridge University Press, Champaign, IL.
Wu, C.B. 2003. Optimal calibration estimators in survey sampling. Biometrika, **90**(4): 937–951. doi:10.1093/biomet/90.4.937.
Wulder, M.A., White, J.C., Nelson, R.F., Næsset, E., Ørka, H.O., Coops, N.C., Hilker, T., Bater, C.W., and Gobakken, T. 2012. Lidar sampling for large-area forest characterization: A review. Remote Sens. Environ. **121**(0): 196–209. doi:10.1016/j.rse.2012.02.001.

Appendix A

Model-dependent estimators of variance

Model-dependent (MD) estimators of variance reflects the uncertainty in an estimated model conditional on a correct model form (Gregoire 1998); it is not an estimator of sampling variance.

The MD estimator of variance of a regression estimator of an AOI mean $\hat{\bar{y}}_i^{\mathrm{rgr}}$ can be obtained from eq. 7.4 on page 73 in Chambers and Clark (2012),

$$\text{(A1)}\qquad \hat{\mathrm{var}}_{\mathrm{MD}}\left(\hat{\bar{y}}_i^{\mathrm{rgr}}\right) = N_i^{-2}\left\{\hat{\sigma}_{ei}^2(N_i - n_i) + \boldsymbol{\Delta}'\hat{\boldsymbol{\Sigma}}_{\hat{\beta}}\boldsymbol{\Delta}\right\}$$

where N_i is the population size of the ith AOI (here number of forested units, viz. pixels), and $\hat{\sigma}_{ei}^2$ is the least-squares estimate of the residual variance, $\boldsymbol{\Delta}$ is a column vector of p differences (p is the number of parameters in a chosen regression model), and $\hat{\boldsymbol{\Sigma}}_{\hat{\beta}}$ is the least-squares estimate of the variance–covariance matrix of the model parameters ($\boldsymbol{\beta}$). The transpose of $\boldsymbol{\Delta}$ is denoted by $\boldsymbol{\Delta}'$.

For model 10 selected for the Swiss FDs we get $\boldsymbol{\Delta}' = (N_i - n_i, N_i\bar{x}_i - n_i\hat{\bar{x}})$, while for model (1) selected for the Vestfold KOMS we get,

$$\text{(A2)}\qquad \boldsymbol{\Delta}' = \left(N_i - n_i, N_i\bar{x}_i - n_i\hat{\bar{x}}, N_i\overline{(x_i^2)} - n_i\overline{(\hat{x}_i^2)}\right).$$

From eq. A1 the MD estimator of relative accuracy was computed as,

$$\text{(A3)}\qquad se_{\mathrm{MD}}\left(\hat{\bar{y}}_i^{\mathrm{rgr}}\right)\% = 100 \times \sqrt{\hat{\mathrm{var}}_{\mathrm{MD}}\left(\hat{\bar{y}}_i^{\mathrm{rgr}}\right)} \times \left(\hat{\bar{y}}_i^{\mathrm{rgr}}\right)^{-1}.$$

Direct predictions of area means and conditional mean squared errors

The AOI direct MB estimators (cEBLUP, or "cond" for short) applicable to models 11 to 13 are (Chambers and Clark 2012, eq. 15.20),

$$\text{(A4)}\qquad \hat{\bar{y}}_i^{\mathrm{cond}} = \left(\sum_{j=1}^{n_i}\hat{w}_{ij}^{\mathrm{EBLUP}}y_{ij}\right) \times \left(\sum_{j=1}^{n_i}\hat{w}_{ij}^{\mathrm{EBLUP}}\right)^{-1}$$

where $\hat{w}_{ij}^{\mathrm{EBLUP}}$ is a fixed EBLUP weight defined in eq. 15.19 of Chambers and Clark (2012).

A robust variance estimator for $\hat{\bar{y}}_i^{\mathrm{cond}}$ is given in eq. 15.24 by Chambers and Clark (2012),

$$\text{(A5)}\qquad \hat{\mathrm{var}}\left(\hat{\bar{y}}_i^{\mathrm{cond}}\right) = N_i^{-2}\sum_{j=1}^{n_i}\hat{\delta}_{ij}(y_{ij} - \tilde{y}_{ij})^2$$

where $\tilde{y}_{ij}$ is an approximately unbiased (in large samples) estimator of the expectation of y_{ij} in area i conditional on the auxiliary variables and the small area effects ($\mathbf{u}_i$), and $\hat{\delta}_{ij}$ is a weight defined in (A7). For any linear model with the intercept included in $\mathbf{x}$, the estimator of $\tilde{y}_{ij}$ can be written as (Chambers and Clark 2012, eq. 15.22)

$$\text{(A6)}\qquad \tilde{y}_{ij} = \mathbf{x}_{ij}'\hat{\boldsymbol{\beta}}^{\mathrm{EBLUE}} + \mathbf{g}_{ij}(\mathbf{G}_i'\mathbf{G}_i)^{-1}\mathbf{G}_i'\left(\mathbf{y}_i - \overline{\mathbf{x}}_i'\hat{\boldsymbol{\beta}}^{\mathrm{EBLUE}}\right)$$

where $\mathbf{g}_{ij}$ is the subset of $\mathbf{x}_{ij}$ with a random area effect, $\mathbf{G}_i$ $\mathbf{G}_i$ is the corresponding area-level subset of $\mathbf{x}_i$, and $\hat{\boldsymbol{\beta}}^{\mathrm{EBLUE}}$ is the empirical best linear unbiased estimator of the fixed-effects regression coefficients (Chambers and Clark 2012, p. 166).

$$\text{(A7)}\qquad \hat{\delta}_{ij} = \left\{(1 - \hat{\phi}_{ijij})^2 + \sum_{k=1}^{n_{\mathrm{AOI}}}\sum_{l=1}^{n_k}\hat{\phi}_{ijkl}^2(1 - \delta_{k=i,l=j})\right\}^{-1} \times \left\{\hat{\theta}_{ij}^2 + (N_i - n_i)n_i\right\}$$

where $\hat{\phi}_{ijkl}$, $k = 1, \ldots, n_k$, $l = 1, \ldots, n_l$ is a set of fixed weights determined so that $\tilde{y}_{ij} = \sum_{k=1}^{n_{\mathrm{AOI}}}\sum_{l=1}^{n_k}\hat{\phi}_{ijkl}y_{kl}$, and $\hat{\theta}_{ij}$ is given below eq. 15.23 in Chambers and Clark (2012). The coefficients $\hat{\phi}_{ijkl}$ are found by expressing $\tilde{y}_{ij}$ as a linear sum of y_{ij}, $i = 1, \ldots, n_{dom}$, $j = 1, \ldots, n_i$.

The conditional mean squared error of $\hat{\bar{y}}_i^{\mathrm{cond}}$ is obtained by adding the square of the estimated bias of $\hat{\bar{y}}_i^{\mathrm{cond}}$ to $\hat{\mathrm{var}}(\hat{\bar{y}}_i^{\mathrm{cond}})$. A conservative estimator of this mean squared error is (Chambers and Clark 2012, eq. 15.26),

$$\text{(A8)}\qquad \mathrm{MSE}\left(\hat{\bar{y}}_i^{\mathrm{cond}}\right) = \hat{\mathrm{var}}\left(\hat{\bar{y}}_i^{\mathrm{cond}}\right) + \hat{\mathrm{B}}\mathrm{ias}^2\left(\hat{\bar{y}}_i^{\mathrm{cond}}\right).$$

The estimator of bias in (A8) is,

$$\text{(A9)}\qquad \hat{\mathrm{B}}\mathrm{ias}\left(\hat{\bar{y}}_i^{\mathrm{cond}}\right) = \left(\sum_{j=1}^{n_i}\hat{w}_{ij}^{\mathrm{EBLUP}}\right)^{-1}\sum_{j=1}^{n_i}\hat{w}_{ij}^{\mathrm{EBLUP}}\tilde{y}_{ij} - N_i^{-1}\sum_{j=1}^{N_i}\tilde{y}_{ij}$$

where $\hat{w}_{ij}^{\mathrm{EBLUP}}$, $j = 1, \ldots, n_i$ is a set of fixed constants satisfying.

$$\text{(A10)}\qquad \hat{\bar{y}}_i^{\mathrm{cond}} = \left(\sum_{j=1}^{n_i}\hat{w}_{ij}^{\mathrm{EBLUP}}\right)^{-1}\sum_{j=1}^{n_i}\hat{w}_{ij}^{\mathrm{EBLUP}}y_{ij}.$$

The weights are specified in eq. 15.19 in Chambers and Clark (2012).

The relative RMSE of $\hat{\bar{y}}_i^{\mathrm{cond}}$ i.e., $\mathrm{RMSE}(\hat{\bar{y}}_i^{\mathrm{cond}})\%$ was computed as,

$$\text{(A11)}\qquad \mathrm{MSE}\left(\hat{\bar{y}}_i^{\mathrm{cond}}\right)^{0.5}\left(\hat{\bar{y}}_i^{\mathrm{cond}}\right)^{-1} \times 100.$$

All EBLUP and cEBLUP results were computed with the Mathematica® version 9.0.1® software (Wolfram 1999).

Appendix B

Spatial correlation of random area effects

The random area effects in models 11, 12, and 13 (Table 1) are assumed uncorrelated. A more natural assumption is to expect a correlation between "similar" AOIs (Chandra et al. 2007; Cressie 1993, p. 383; Molina et al. 2009; Petrucci et al. 2005; Pratesi and Salvati 2008). A positive spatial correlation can be exploited in SAE to improve accuracy (Chandra et al. 2007; Molina et al. 2009; Pfeffermann 2013; Pratesi and Salvati 2008).

A priori, we conjectured that a spatial correlation of random area effects would not reflect a continuous stationary spatial process (Freeman and Moisen 2007) due to complex trends in topography, climate, and forest structures in the two study populations.

We therefore assume a correlation restricted to pairs of AOIs sharing a border. This restriction invites an area-level Fay–Herriot (FH) model with correlated area effects (Chandra et al. 2007; Molina et al. 2009; Petrucci et al. 2005; Pratesi and Salvati 2008).

To gauge the strength of a possible spatial correlation among pairs of neighbouring AOIs, we computed Moran's I statistic (Cliff and Ord 1981, p. 42–47). Moran's I statistic gives the average correlation between random area effects in AOIs sharing a border. Moran's I was computed from standardized residuals from country-specific model 1 regressions and a proximity matrix **W** with element $w_{ij} = 1$ if the ith and jth AOI shares a border, and otherwise 0.

When Moran's I suggested a significant positive spatial correlation, we computed FD (KOM) spatial EBLUP (sEBLUP) means and estimates of mean squared errors using a Fay–Herriot model (FH) with: (*i*) the area-level sample mean of y as the dependent variable, (*ii*) the known area mean of **X** as a vector of fixed-effects predictors, (*iii*) a spatially correlated random effect (Rao 2003, p. 168), and (*iv*) a random and independent residual error. A row-normalized version of **W** was used in these calculations. Details are in the R-package "sae" (Molina et al. 2009). Note, the FH model is not part of the model selection process or model averaging. Results with sEBLUP provide grounds for comparisons with RMSEs of unit-level models 12 and 13.