Article

A double-sampling extension of the German National Forest Inventory for design-based small area estimation on forest district levels

Andreas Hill [1,*], Daniel Mandallaz [1] and Joachim Langshausen [2]

[1] Department of Environmental Systems Science, ETH Zurich, 8092 Zurich, Switzerland; mdaniel@retired.ethz.ch (D.M.)
[2] State Forest Service Rhineland-Palatinate, 56281 Emmelshausen, Germany; Joachim.Langshausen@wald-rlp.de
* Correspondence: andreas.hill@usys.ethz.ch; Tel.: +x-xxx-xxx-xxxx

Academic Editor: name
Version 9th May 2018 submitted to Remote Sens.

Abstract: The German National Forest Inventory consists of a systematic grid of permanent sample plots and provides a reliable evidence-based assessment of the state and the development of Germany's forests on national and federal state level in a 10 year interval. However, the data have yet been scarcely used for estimation on smaller management levels such as forest districts due to insufficient sample sizes within the area of interests and the implied large estimation errors. In this study, we present a double-sampling extension to the existing German National Forest Inventory (NFI) that allows for the application of recently developed design-based small area regression estimators. We illustrate the implementation of the estimation procedure and evaluate its potential by the example of timber volume estimation on two small scale management levels (45 and 405 forest district units respectively) in the federal German state of Rhineland-Palatinate. An airborne laserscanning (ALS) derived canopy height model and a tree species classification map based on satellite data were used as auxiliary data in an ordinary least square regression model to produce the timber volume predictions. The results support that the suggested double-sampling procedure can substantially increase estimation precision on both management levels: the two-phase estimators were able to reduce the variance of the SRS estimator by 43% and 25% on average for the two management levels respectively.

Keywords: National forest inventory, small area estimation, forest districts, double sampling for regression within strata, cluster sampling, canopy height model, tree species classification

1. Introduction

The German National Forest Inventory (NFI) provides reliable evidence-based and accurate information of the current state and the development of Germany's forest over time. The NFI thereby has the responsibility to satisfy various information needs including reporting to public and state forestry administrations, wood-based industries and the public on the national level, as well as to the Food and Agriculture Organization of the United Nations (FAO) and to the United Nations Framework Convention on Climate Change (UNFCC) on the international level [1]. The current design of the German NFI rests solely upon a terrestrial cluster inventory that is carried out at sample locations systematically distributed over the entire forest state area of Germany. In order to cover a large area of 114'191 ha [2], the sample size has been specifically chosen to satisfy high estimation accuracies for forest attributes on the national and federal state levels. However, sample sizes often drop dramatically when entering spatial units below the federal state level. This is particularly true for forest management levels such as forest districts for which the estimation uncertainties turn out to be unacceptably large due to the very limited number of sample plots within these units. For this reason, the German NFI data have not yet been extensively incorporated into operational planning on forest district management levels. In most German federal states, management strategies are thus still based on expert judgments from time-consuming standwise inventories (SFI), which are prone to systematic deviations [3] and do not provide any measure of uncertainty.

Some German federal states, such as Lower Saxony, have approached this problem by establishing a regional Forest District Inventory (FDI) with a much higher sampling density than used by the NFI in order to scientifically base their regional management strategies on quantitative and accurate information [4]. However, such FDIs are cost-intensive and, facing increasing restrictions in budget and staff resources, there has been a need for more cost-efficient inventory methods [5]. One method which has proven to be efficient is double- or two-phase sampling [6–9]. Double-sampling incorporates less expensive auxiliary information and can be used to either increase estimation precision under a fixed terrestrial sample size, or maintain estimation precision under reduced terrestrial sample size. Double-sampling procedures have already been used for stratification in the FDI of Lower Saxony [10], and Grafström *et al.* [11] illustrated how to use the auxiliary information to determine optimised balanced terrestrial sample designs. Recent studies have extended double-sampling to triple-sampling estimation methods using auxiliary information derived at two different sampling intensities. An example can be found in von Lüpke *et al.* [12] who illustrated an extension of the existing two-phase FDI of Lower Saxony to a three-phase design that uses updates of past inventory data as additional auxiliary information and allows for a significant reduction of the terrestrial sample size in intermediate inventories. Another example is Massey *et al.* [13] who developed a triple-sampling extension based on the ideas of Mandallaz [14] for the Swiss NFI that can significantly reduce the increase in estimation uncertainty caused by the new annual inventory design.

Two-phase and three-phase samplings techniques have also been applied to small area estimation (SAE). SAE techniques address the situation where the number of samples within a subunit, or small area (SA), of the entire sampling frame is too small to provide reliable estimates for that unit. A broad range of SA estimators used in forest inventories [8] originally comes from official statistics. One such method that is commonly applied is known as indirect estimation [15], where statistical models are used to convert auxiliary information into predictions of the target variable that is rarely or not observed in the small area. These models are trained using data from outside the small area in order to "borrow strength" from areas where information is available. Of numerous applications of SAE in forestry [16–19], most use unit-level models, i.e. the inventory plot is the unit of the response variable in the training data used for the model fit. Such unit-level models have been intensively investigated for timber volume estimation using various remote sensing auxiliary data [20,21]. Other studies have investigated area-level models, where the auxiliary information is only provided on the SA level [22]. Some studies have illustrated that even NFI data derived under low sampling densities can still be used to provide acceptable precision of small area estimates on much smaller management levels. One example is Breidenbach and Astrup [16] who used data from the Norwegian NFI to make small area estimation for standing timber volume for 14 municipalities where the number of NFI samples within these areas were between 1 and 35. The estimation errors under the applied model-dependent and design-based small area estimators turned to be markedly smaller than under the standard one-phase estimator. Another example is Magnussen *et al.* [23] who recently used the Swiss NFI data to estimate timber volume within 108 Swiss forest districts with sample sizes between 9 and 206. Similar studies using German NFI data for small area estimation have been lacking.

The aim of this study was to investigate whether the German NFI data can provide acceptable estimation precision on two forest district levels using the latest small area estimation procedures. We therefore conducted a study in the German federal state Rhineland-Palatinate where we extended the German NFI to a double-sampling design and applied three types of design-based small area regression estimators in order to derive point and variance estimates of mean standing timber volume for 45 and 405 forest districts respectively. The SA-estimators we considered were the *pseudo-small*, *extended pseudo-synthetic* and the *pseudo-synthetic* design-based small area estimator suggested by Mandallaz [24] and Mandallaz *et al.* [19]. Auxiliary data consisted of a canopy height model (CHM) obtained from a countrywide airborne laser scanning (ALS) and a tree species classification map to be used for regression within tree species strata. The estimation precisions were compared to those obtained by the standard one-phase estimator for cluster sampling under simple random sampling. The chosen double-sampling estimators were selected for several reasons: **(i)** the design-based framework relaxes dependencies on the regression model assumptions which seemed appropriate facing severe quality restrictions in the ALS data; **(ii)** the estimators can be used with *non-exhaustive*, i.e. non wall-to-wall, auxiliary information; **(iii)** all estimators are explicitly formulated for cluster sampling which has not yet been the case for frequently used model-dependent estimators; and **(iv)** the asymptotically unbiased g-weight variance accounts for estimating the regression coefficients on the same sample used for estimation (*internal model approach*) and is also robust under heteroscedasticity of the model residuals. The results from this study were considered to provide valuable information whether the suggested procedure might be a cost-saving alternative to a regional FDI.

2. Terrestrial sampling design of the German NFI

The German NFI is a periodic inventory that is carried out every 10 years over the entire forest area of Germany. The most recent inventory (BWI3) was conducted in 2011 and 2012. While information was originally gathered on a systematic 4x4 km grid, some federal states such as Rhineland-Palatinate have switched to a densified 2x2 km grid. The German NFI uses a cluster sampling design, which means that a sample unit consists of at most four sample locations (also referred to as *sample plots*) that are arranged in a square, called *cluster*, with a side length of 150 metres. The number of plots per cluster can vary between 1 and 4 depending on forest/non-forest decisions by the field crews on the individual plot level [25]. In the field survey of the BWI3, sample trees for timber volume estimation are selected according to the angle count sampling technique [26], using a basal area factor (BAF) of 4 that is respectively adjusted for sample trees at the forest boundary by a geometric intersection of the boundary transect with the individual tree's inclusion circle [25]. A further inventory threshold for a tree to be recorded is a diameter at breast height (DBH) of at least 7 cm. For each sample tree that is selected by this procedure, the DBH, the absolute tree height, the tree diameter at 7 m ($D7$) and the tree species is measured and used to estimate the volume at the tree level. These volume estimates are based on the application of tree species specific taper curves that are adjusted to the set of diameters and corresponding height measurements taken from the respective sample tree [27].

3. Double sampling in the infinite population approach

3.0.1. One- and Two-Phase Sampling in the Infinite Population Approach

The estimators used in this study have been proposed by [19,24] and derive their mathematical properties under the so-called infinite population approach. Therefore, we shall first provide a short introduction into this general estimation framework. We start by assuming that the population P of trees $i \in 1, 2, ..., N$ within a forest of interest F is exactly defined, and each tree i has a response variable Y_i (e.g. its timber volume) that can be used to define the population mean Y (e.g., the average timber volume per unit area) over F. Since a full census of all tree population individuals is almost never feasible, Y has to be estimated based on a sample. In the infinite population approach this sample is a set of points or locations x distributed independently and uniformly over the set of all possible points in F. Each point x has an associated local density $Y(x)$ (e.g., the timber volume per unit area) whose spatial distribution is given by a fixed (i.e. non stochastic) piecewise constant function. The population mean Y is mathematically equivalent to the integral of the local density function surface divided by the surface area of F, $\lambda(F)$, i.e. $Y = \frac{1}{N}\sum_{i=1}^{N} Y_i = \frac{1}{\lambda(F)}\int_F Y(x)dx$, and thus the population mean Y corresponds to a spatial mean. Since the actual local density function is unobserved in its entirety, one estimates Y by taking a sample s_2 consisting of n_2 points and measuring each of their respective local densities. This sampling procedure is often referred to as *one-phase sampling* (OPS) and s_2 is referred to as the terrestrial inventory. In contrast to the one-phase approach, *two-phase* or *double-sampling* procedures use information from two nested samples (phases). Practically speaking, the terrestrial inventory s_2 is embedded in a large phase s_1 comprising n_1 sample locations that each provide a set of explanatory variables described by the column vector $\mathbf{Z}(x) = (z(x)_1, z(x)_2, ..., z(x)_p)^\top$ at each point $x \in s_1$. These explanatory variables are derived from auxiliary information that is available in high quantity within the forest F. For every $x \in s_1$, $\mathbf{Z}(x)$ is transformed into a prediction $\hat{Y}(x)$ of $Y(x)$ using the choice of some prediction model. The basic idea of this method is to boost the sample size by providing a large sample of less precise but cheaper predictions of $Y(x)$ in s_1 and to correct any possible model bias, i.e., $\mathbb{E}(Y(x) - \hat{Y}(x))$, using the subsample of terrestrial inventory units where the value of $Y(x)$ is observed. In this context, it is also important to note that the response and auxiliary variables are assumed to be error-free and the resulting errors for the point estimates reflect only the uncertainty due to sampling.

4. Estimators

4.1. Design-based one-phase estimator for cluster sampling (SRS)

The one-phase estimator for cluster sampling (SRS) constitutes the *status quo* that is currently applied under the existing one-phase sampling design of the German NFI in order to obtain point and variance estimates for the mean timber volume of a given estimation unit. In order to provide all estimators in the infinite population framework and ensure a consistent terminology with the two-phase estimators in Section 4.2, we will introduce the SRS estimator that is applied in the BWI3 algorithms [28] in the form given in Mandallaz [9], Mandallaz *et al.* [29].

In order to calculate the local density $Y_c(x)$ at the cluster level, a cluster is defined as consisting of M sample locations (in the BWI3, we have $M = 4$) where $M - 1$ sample locations $x_2, ..., x_M$ are created close to the cluster origin x_1 by adding a fixed set of spatial vectors $e_2, ..., e_M$ to x_1. The actual number of plots per cluster, $M(x)$, is a random variable due to the uniform distribution of x_l ($l = 1, ..., M$) in the forest F and to the forest/non-forest decision for each sample location x_l:

$$M(x) = \sum_{l=1}^{M} I_F(x_l) \quad \text{where} \quad I_F(x_l) = \begin{cases} 1 \text{ if } x_l \in F \\ 0 \text{ if } x_l \notin F \end{cases} \tag{1}$$

The local density on cluster level $Y_c(x)$, which is in our case the timber volume per hectare, is then defined as the average of the individual sample plot densities $Y(x_l)$:

$$Y_c(x) = \frac{\sum_{l=1}^{M} I_F(x_l) Y(x_l)}{M(x)} \tag{2}$$

The local density $Y(x_l)$ on individual sample plot level was calculated according to the description in Mandallaz [9], which can be rewritten for angle-count sampling technique applied in the BWI3. The general form of $Y(x)$ in Mandallaz [9] is given as the Horwitz-Thompson estimator

$$Y(x_l) = \sum_{i \in s_2(x_l)} \frac{Y_i}{\pi_i \lambda(F)} \tag{3}$$

where Y_i is in our case the timber volume of the tree i recorded at sample location x in m^3. Each tree has an inclusion probability π_i that is well defined as the proportion of its inclusion circle area $\lambda(K_i)$ within the forest area $\lambda(F)$, i.e. via their geometric intersection:

$$\pi_i = \frac{\lambda(K_i \cap F)}{\lambda(F)} \tag{4}$$

The radius R_i of the tree's inclusion circle K_i is given by $R_i = DBH_i / cf_{i,corr}$ (also referred to as *limiting distance*), where $cf_{i,corr}$ is the original counting factor cf corrected for potential boundary effects at the forest border. In case of angle-count sampling, we can rewrite π_i as

$$\pi_i = \frac{G_i}{cf_{i,corr}\lambda(F)} \tag{5}$$

since the intersection area $\lambda(K_i \cap F)/\lambda(F)$ can be expressed using the trees basal area G_i (in m^2) and the corrected counting factor:

$$\lambda(K_i \cap F) = \frac{G_i}{cf_{i,corr}} \quad \text{where} \quad cf_{i,corr} = cf \frac{\lambda(K_i)}{\lambda(K_i \cap F)} \tag{6}$$

Eq. 5 in Eq. 3 yields the rewritten form of $Y(x_l)$ for angle count sampling that conforms to the definition used in the BWI3 algorithms [28]:

$$Y(x_l) = \sum_{i \in s_2(x_l)} \frac{cf_{i,corr} Y_i}{G_i} = \sum_{i \in s_2(x_l)} nha_i Y_i \tag{7}$$

where nha_i is the number of trees per hectare represented by tree i. The local densities on cluster level can then be used to derive the estimated spatial mean $\hat{\hat{Y}}_c$ and its estimated variance $\hat{\mathbb{V}}(\hat{\hat{Y}}_c)$ for any given spatial unit for which $n_2 \geq 2$ (n_2 denoting the number of clusters):

$$\hat{\hat{Y}}_c = \frac{\sum_{x \in s_2} M(x) Y_c(x)}{\sum_{x \in s_2} M(x)} \tag{8a}$$

$$\hat{\mathbb{V}}(\hat{\hat{Y}}_c) = \frac{1}{n_2(n_2-1)} \sum_{x \in s_2} \left(\frac{M(x)}{\bar{M}_2} \right)^2 (Y_c(x) - \hat{\hat{Y}}_c)^2 \tag{8b}$$

with $\bar{M}_2 = \frac{\sum_{x \in s_2} M(x)}{n_2}$.

4.2. Design-based small area regression estimators for cluster sampling

All three considered small area estimators use ordinary least square (OLS) regression models to produce predictions of the local density $Y_c(x)$ directly on the cluster level c. We consider the internal model approach, where the estimators take into account that the regression coefficients on the cluster level were fitted using the same sample used for estimation. To apply this to small area estimation, the vector of estimated regression coefficients on the cluster level is found by "borrowing strength" from the entire terrestrial sample s_2 of the current inventory:

$$\hat{\boldsymbol{\beta}}_{c,s_2} = \boldsymbol{A}_{c,s_2}^{-1} \left(\frac{1}{n_2} \sum_{x \in s_2} M(x) Y_c(x) \mathbf{Z}_c(x) \right) \tag{9a}$$

$$\boldsymbol{A}_{c,s_2} = \frac{1}{n_2} \sum_{x \in s_2} M(x) \mathbf{Z}_c(x) \mathbf{Z}_c^{\top}(x) \tag{9b}$$

$\mathbf{Z}_c(x)$ is the vector of explanatory variables on the cluster level, which is calculated as the weighted average of the explanatory variables $\mathbf{Z}(x_l)$ on the individual plot levels $x_1, ..., x_l$ (Eq.10). The weight $w(x_l)$ is the proportion of the support-area within the forest F used to derive the explanatory variables from the raw auxiliary information.

$$\mathbf{Z}_c(x) = \frac{\sum_{l=1}^{M} I_F(x_l) w(x_l) \mathbf{Z}(x_l)}{\sum_{l=1}^{M} I_F(x_l) w(x_l)} \tag{10}$$

The estimated design-based variance-covariance matrix $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}_{c,s_2}}$ accounts for the fact that the regression model is internal and reflects the sampling variability that occurs when estimating the regression coefficients on the realized sample s_2. It is defined as

$$\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}_{c,s_2}} = \boldsymbol{A}_{c,s_2}^{-1}\Big(\frac{1}{n_2^2}\sum_{x\in s_2} M^2(x)\hat{R}_c^2(x)\mathbf{Z}_c(x)\mathbf{Z}_c^{\top}(x)\Big)\boldsymbol{A}_{c,s_2}^{-1} \tag{11}$$

with

$$\hat{R}_c = Y_c(x) - \mathbf{Z}_c^{\top}(x)\hat{\boldsymbol{\beta}}_{c,s_2} = Y_c(x) - \hat{Y}_c(x) \tag{12}$$

being the empirical model residuals at the cluster level, which by construction of OLS satisfy the important zero mean residual property, i.e. $\frac{\sum_{x\in s_2} M(x)\hat{R}_c(x)}{\sum_{x\in s_2} M(x)} = 0$.

In the following, we will give a short description of each small area estimator and refer to Mandallaz *et al.* [19], Mandallaz [24], Mandallaz *et al.* [29] if the reader requires additional details or proofs. The estimators have also been implemented in the R-package *forestinventory* [30] which was used to compute all estimates in this study.

4.2.1. Pseudo Small Area Estimator (PSMALL)

All point information used for small area estimation is now restricted to that available at the sample locations $s_{1,G}$ or $s_{2,G}$ in the small area G, with exception of $\hat{\boldsymbol{\beta}}_{c,s_2}$ and $\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}_{c,s_2}}$ which are always based on the entire sample s_2. We thus first define the following quantities on the small area level:

$$\hat{\bar{\mathbf{Z}}}_{c,G} = \frac{\sum_{x\in s_{1,G}} M_G(x)\mathbf{Z}_{c,G}(x)}{\sum_{x\in s_{1,G}} M_G(x)} \quad \text{where} \quad \mathbf{Z}_{c,G}(x) = \frac{\sum_{l=1}^{L} I_G(x_l)\mathbf{Z}(x_l)}{M_G(x)} \tag{13a}$$

$$Y_{c,G}(x) = \frac{\sum_{l=1}^{L} I_G(x_l)Y(x_l)}{M_G(x)} \quad \text{and} \quad \hat{Y}_{c,G}(x) = \hat{\bar{\mathbf{Z}}}_{c,G}^{\top}\hat{\boldsymbol{\beta}}_{c,s_2} \tag{13b}$$

$$\hat{\bar{R}}_{2,G} = \frac{\sum_{x\in s_{2,G}} M_G(x)\hat{R}_{c,G}(x)}{\sum_{x\in s_{2,G}} M_G(x)} \quad \text{where} \quad \hat{R}_{c,G}(x) = Y_{c,G}(x) - \hat{Y}_{c,G}(x) \tag{13c}$$

Note that the restriction to G, i.e. $I_G(x_l) = \{0,1\}$, is made on the individual sample plot level x_l, and $M_G(x) = \sum_{l=1}^{L} I_G(x_l)$ thus is the number of sample plots per cluster within the small area. The asymptotically design-unbiased point estimate of *PSMALL* is then defined according to Eq. 14a. The first term estimates the small area population mean of G by applying the globally derived regression coefficients to the small area cluster means of the explanatory variables $\hat{\bar{\mathbf{Z}}}_{c,G}$. The second term then corrects for a potential bias of the regression model predictions in the small area G by adding the mean of the empirical residuals $\hat{\bar{R}}_{2,G}$ in G. This correction is necessary because the zero mean residual property that holds in F is not guaranteed to hold in small area G under this construction.

$$\hat{Y}_{c,G,PSMALL} = \hat{\bar{\mathbf{Z}}}_{c,G}^{\top}\hat{\boldsymbol{\beta}}_{c,s_2} + \hat{\bar{R}}_{2,G} \tag{14a}$$

$$\begin{aligned}\hat{\mathbb{V}}(\hat{Y}_{c,G,PSMALL}) = {} & \hat{\bar{\mathbf{Z}}}_{c,G}^{\top}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}_{c,s_2}}\hat{\bar{\mathbf{Z}}}_{c,G} + \hat{\boldsymbol{\beta}}_{c,s_2}^{\top}\hat{\Sigma}_{\hat{\bar{\mathbf{Z}}}_{c,G}}\hat{\boldsymbol{\beta}}_{c,s_2} \\ & + \frac{1}{n_{2,G}(n_{2,G}-1)}\sum_{x\in s_{2,G}}\Big(\frac{M_G(x)}{\bar{M}_{2,G}}\Big)^2(\hat{R}_{c,G}(x) - \hat{\bar{R}}_{2,G})^2\end{aligned} \tag{14b}$$

with $\bar{M}_{2,G} = \frac{\sum_{x \in s_{2,G}} M_G(x)}{n_{2,G}}$.

The variance-covariance matrix of the auxiliary vector $\hat{\Sigma}_{\hat{\bar{\mathbf{Z}}}_{c,G}}$ is thereby defined as

$$\hat{\Sigma}_{\hat{\bar{\mathbf{Z}}}_{c,G}} = \frac{1}{n_{1,G}(n_{1,G}-1)} \sum_{x \in s_{1,G}} \left(\frac{M_G(x)}{\bar{M}_{1,G}}\right)^2 (\mathbf{Z}_{c,G}(x) - \hat{\bar{\mathbf{Z}}}_{c,G})(\mathbf{Z}_{c,G}(x) - \hat{\bar{\mathbf{Z}}}_{c,G})^\top \tag{15}$$

with $\bar{M}_{1,G} = \frac{\sum_{x \in s_{1,G}} M_G(x)}{n_{1,G}}$.

The estimated design-based variance of $\hat{Y}_{c,G,PSMALL}$ is given by Eq. 14b. Basically, the first term constitutes the variance introduced by the uncertainty in the regression coefficients, whereas the second term expresses the variance caused by estimating the exact auxiliary mean in G using a non-exhaustive sample $s_{1,G}$. The third term is the variance of the model residuals and thus accounts for the inaccuracies of the model predictions. Note that the first term can also be rewritten using g-weights [29, pg.14] which ensures some beneficial calibration of the auxiliary variables to the first-phase sample.

4.2.2. Pseudo Synthetic Estimator (PSYNTH)

The PSYNTH estimator is commonly applied when no terrestrial sample is available within the small area G (i.e. $n_{2,G} = 0$). The point estimate (Eq. 16a) is thus only based on the predictions generated by applying the globally derived regression coefficients to the small area cluster means of the explanatory variables $\hat{\bar{\mathbf{Z}}}_{c,G}$. Note that the bias correction term using the empirical residuals (Eq. 14a) can no longer be applied. The PSYNTH estimator thus has a potential unobservable design-based bias.

$$\hat{Y}_{c,G,PSYNTH} = \hat{\bar{\mathbf{Z}}}_{c,G}^\top \hat{\boldsymbol{\beta}}_{c,s_2} \tag{16a}$$

$$\hat{\mathbb{V}}(\hat{Y}_{c,G,PSYNTH}) = \hat{\bar{\mathbf{Z}}}_{c,G}^\top \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}_{c,s_2}} \hat{\bar{\mathbf{Z}}}_{c,G} + \hat{\boldsymbol{\beta}}_{c,s_2}^\top \hat{\Sigma}_{\hat{\bar{\mathbf{Z}}}_{c,G}} \hat{\boldsymbol{\beta}}_{c,s_2} \tag{16b}$$

The contribution to the variance by the model residuals in small area G can also no longer be considered (Eq. 16b). As a result, the synthetic estimator will usually have a smaller variance than estimators that consider the model residuals, but at the cost of a potential bias. Note that the PSYNTH estimator is still design-based, but one purely has to rely on the validity of the regression model within the small area as it is the case in the model-dependent framework.

4.2.3. Extended Pseudo Synthetic Estimator (EXTPSYNTH)

The EXTPSYNTH estimator (Eq. 17) has been proposed by [24] as a transformed version of the PSMALL estimator that has the form of the PSYNTH estimator but remains asymptotically design unbiased. It has the advantage that the mean of the empirical model residuals of the OLS regression model for the entire area F and the small area G are by construction both zero at the same time, i.e. $\hat{\bar{R}}_c = \hat{\bar{R}}_{c,G} = 0$. This is realized by *extending* the auxiliary vector $\mathbf{Z}_c(x)$ by the indicator variable $I_{c,G}$ which takes the value 1 if the entire cluster lies within the small area G and 0 if the entire cluster is outside G, i.e. $I_{c,G}(x) = \frac{M_G(x)}{M(x)}$. The extended auxiliary vector thus becomes $\mathbb{Z}_c^\top(x) = (\mathbf{Z}_c^\top(x), I_{c,G}(x))$ and the new regression coefficient using $\mathbb{Z}_c(x)$ instead of $\mathbf{Z}_c(x)$ in Eq. 9 is denoted as $\hat{\boldsymbol{\theta}}_{s_2}$. All remaining components are calculated by plugging in $\mathbb{Z}_c(x)$ in Eq. 13. A decomposition of $\hat{\boldsymbol{\theta}}_{s_2}$ reveals that the residual correction term is now included in the regression coefficient $\hat{\boldsymbol{\theta}}_{s_2}$.

$$\hat{Y}_{c,G,EXTPSYNTH} = \hat{\bar{\mathbb{Z}}}_{c,G}^{\top}\hat{\boldsymbol{\theta}}_{c,s_2} \tag{17a}$$

$$\hat{\mathbb{V}}(\hat{Y}_{c,G,EXTPSYNTH}) = \hat{\bar{\mathbb{Z}}}_{c,G}^{\top}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\theta}}_{c,s_2}}\hat{\bar{\mathbb{Z}}}_{c,G} + \hat{\boldsymbol{\theta}}_{c,s_2}^{\top}\hat{\boldsymbol{\Sigma}}_{\hat{\bar{\mathbb{Z}}}_{c,G}}\hat{\boldsymbol{\theta}}_{c,s_2} \tag{17b}$$

However, it is important to note that $\hat{\bar{R}}_{c,G} = 0$ under the extended regression model only holds if the sample plots $x_1, ..., x_l$ of a cluster are *all* either inside our outside the small area, i.e. $M_G(x) \equiv M(x)$, and thus $I_{c,G}(x) = \frac{M_G(x)}{M(x)}$ can only take the values 1 or 0. Mandallaz *et al.* [29] assumed that the effects on the estimates should be negligible as the number of occasions where $M_G(x) < M(x)$ was considered to be small in practical implementations. It was thus a further objective of this study to investigate the actual occurrences and effects of this phenomenon by comparing the estimates of EXTPSYNTH to those of PSMALL.

4.3. Measures of estimation accuracy

The estimation precision was quantified by the estimation error, which is the ratio of the standard error and the point estimate:

$$error_{[\%]} = \frac{\sqrt{\hat{\mathbb{V}}(\hat{Y})}}{\hat{Y}} * 100 \tag{18}$$

We further calculated the 95% confidence interval for each estimate for visualization purposes. The confidence intervals can also be used heuristically for hypothesis testing to determine whether the point estimates of the three estimators for a given small area are statistically different. The confidence intervals for the SRS estimator can be obtained as:

$$CI_{1-\alpha}(\hat{\hat{Y}}_c) = \hat{\hat{Y}}_c \pm t_{n_2-1,1-\frac{\alpha}{2}}\sqrt{\hat{\mathbb{V}}(\hat{\hat{Y}}_c)} \tag{19}$$

The confidence intervals for the PSMALL and EXTPSYNTH estimates are calculated as:

$$CI_{1-\alpha}(\hat{Y}_{c,G,EXTPSYNTH}) = \hat{Y}_{c,G,EXTPSYNTH} \pm t_{n_{2,G}-1,1-\frac{\alpha}{2}}\sqrt{\hat{\mathbb{V}}(\hat{Y}_{c,G,EXTPSYNTH})} \tag{20a}$$

$$CI_{1-\alpha}(\hat{Y}_{c,G,PSMALL}) = \hat{Y}_{c,G,PSMALL} \pm t_{n_{2,G}-1,1-\frac{\alpha}{2}}\sqrt{\hat{\mathbb{V}}(\hat{Y}_{c,G,PSMALL})} \tag{20b}$$

For the PSYNTH estimates, the confidence intervals are

$$CI_{1-\alpha}(\hat{Y}_{c,G,PSYNTH}) = \hat{Y}_{c,G,PSYNTH} \pm t_{n_2-p,1-\frac{\alpha}{2}}\sqrt{\hat{\mathbb{V}}(\hat{Y}_{c,G,PSYNTH})} \tag{21}$$

with p being the number of parameters used in the regression model including the intercept term.

In order to address the potential benefits of the small area estimators compared with the SRS approach, we calculated the *relative efficiency* (Eq. 22) which can be interpreted as the relative sample size under SRS needed to achieve the variance under the double-sampling (DS) estimators.

$$rel.eff = \frac{\hat{\mathbb{V}}(\hat{Y}_{SRS})}{\hat{\mathbb{V}}(\hat{Y}_{DS})} \tag{22}$$

5. Case study

5.1. Study area and small area units

The German federal state Rhineland-Palatinate (*RLP*) is located in the western part of Germany and borders Luxembourg, France and Belgium. With 42.3% (appr. 8400 km^2) of the entire state area (19850 km^2) covered by forest, RLP is one of the two states with the highest forest coverage among all federal states of Germany [2]. The forests of RLP are further characterised by a pronounced diversity in bioclimatic growing conditions that have strong influence on the local growth dynamics as well as tree species composition [31] and are further characterised by large variety of forest structures ranging from characteristic oak coppices (Moselle valley), pure spruce, beech and scots pine forests (i.a. Hunsrück and Palatinate forest) up to mixed forests comprising variable proportions of oak, larch, spruce, Scots pine and beech. Around 82% of the forest area in RLP are mixed forest stands and 69% of the forest area exhibit a multi-layered vertical structure. The forest area of RLP are divided into 3 ownership classes, i.e. state forest (27%), communal forest (46%) and privately owned forest (27%). The forest service of RLP has the legal mandate to sustainably manage the state and communal forest area (73% of the entire forest area), including forest planning, harvesting and the sale of wood [32]. For this reason, the entire forest area has been spatially organised in 3 main hierarchical management units (Figure 1). On the upper level, RLP has been divided into 45 Forstämter (*FA*), which are further divided into a total number of 405 Forstreviere (*FR*). The next level are the forest stands (104'184 in total) for which expert judgements are conducted by SFIs in a 5 to 10 year period in order to set up management strategies for the upcoming 10 years. The FAs and FRs constituted the SA units for which design-based small area estimations of the mean standing timber volume were calculated by incorporating the available terrestrial inventory data of the BWI3 in the estimators described in Section 4. The average area of the SA units was 43'777 ha on the FA-level, and 4624 ha on the FR level.

5.2. Terrestrial sample

Rhineland-Palatinate (RLP) is covered by a 2x2 km inventory grid of the German NFI. In the last inventory (BWI3) conducted in the year 2013, timber volume information was derived for 2810 clusters (8092 plots) in the field survey. The local timber volume density on the plot and cluster level for this sample was consequently calculated according to Section 4.1. In the framework of this survey, the plot center coordinates were re-measured with the differential global satellite navigation system (DGPS) technique. Knowledge about the exact plot positions were considered crucial to provide optimal comparability between the terrestrial observations and the information derived from the auxiliary information. A comparison of the DGPS coordinates with the so-far used target coordinates revealed that 90% of all horizontal deviations lay in the range of 25 meters. A detailed analysis of horizontal DGPS errors in RLP by Lamprecht *et al.* [33] indicated that 80% of the plots should not exceed horizontal DGPS errors of 8 meters. For 162 plots, the DGPS coordinates were replaced by their target coordinates due to missingness or implausible values. The terrestrial sample size $n_{2,G}$ within the FA units was 46 clusters on average and ranged between 11 and 64. Within the FR units, $n_{2,G}$ was considerably smaller with an average of 5 clusters and a range between 0 and 13.

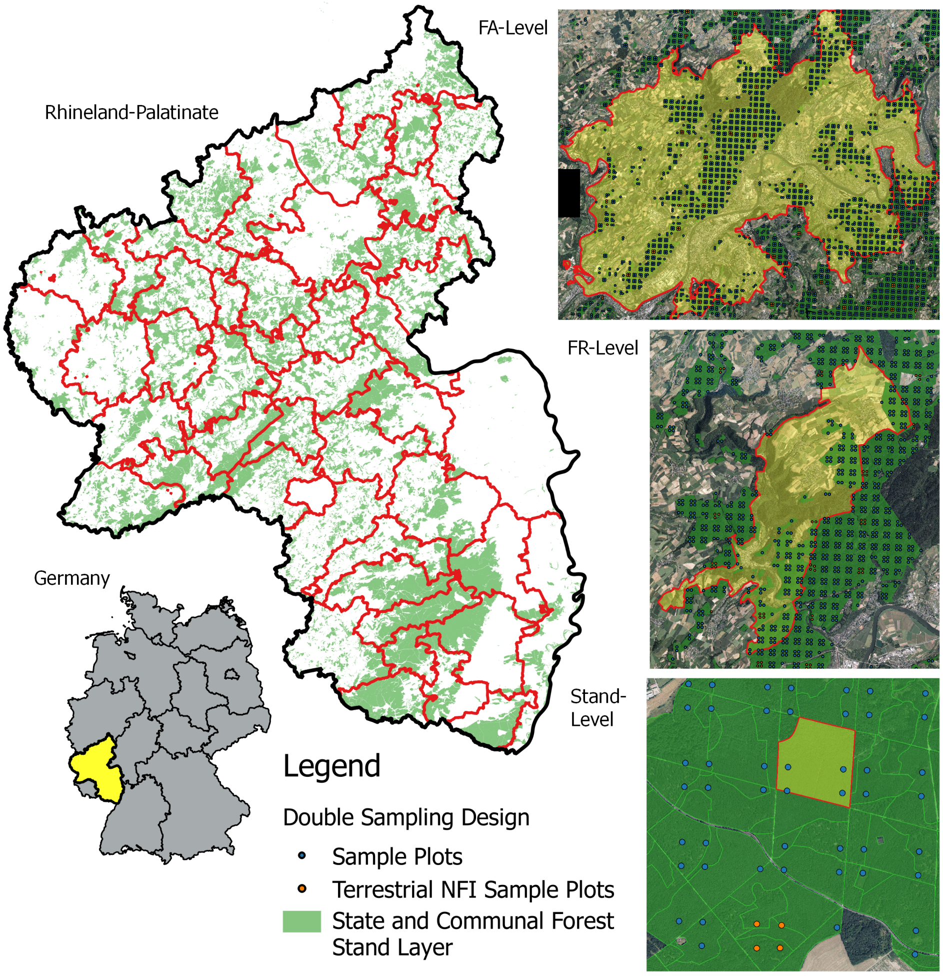

Figure 1. *Left*: Study area with delineated FA forest management units. *Right*: Example for each of the three management units (from top to bottom): FA, FR and forest stand unit overlayed with the extended double-sampling cluster design. *Green*: Forest stand polygon layer defining the forest area of this study.

5.3. Extension to double-sampling design

In order to apply the small area estimators (Section 4.2), the existing NFI design was extended to a double-sampling design by densifying the existing systematic 2x2 km grid to a grid size of 500x500 m that constituted the large first phase s_1 in accordance to Section 3 (Figure 1, *right*). The existing terrestrial phase s_2 was consequently integrated by replacing the target coordinates of the respective s_1 clusters by the terrestrially measured DGPS coordinates. For our study, we restricted the sampling frame to the communal and state forest. The forest/non-forest decision for each plot was thereby made by a spatial intersection of the plot center coordinates with a polygon layer of the communal and state forest stands provided by the forest service. Using this stand layer provided the advantage to consistently apply the same forest/non-forest definition to the entire sample s_1 in order to decide about excluding or including a plot in the sampling frame. The terrestrial sample size n_2 was thus reduced to 2055 clusters (5791 plots). Table 1 provides a short descriptive summary about the volume densities and the main attributes of the NFI plots located in the state and communal forest sampling frame. The densification led to an average sample size $n_{1,G}$ of 759 clusters (range: 246 – 1022) in the FA units, and 88 clusters (range: 1 – 194) in the FR units.

Table 1. Descriptive statistics of the forest observed on NFI sample plots located within communal and state forest area (n=5791).

Variable	Mean	SD	Maximum
Timber Volume (m^3/ha)	300.86	195.55	1375.31
Mean DBH (mm)	354.90	137.22	1123.20
Mean height (dm)	239.60	72.43	497.43
Mean stem density per hectare	101.00	114.01	1010.31

5.4. Auxiliary data

5.4.1. LiDAR canopy height model

A prerequisite for the application of the suggested two-phase small area estimators is the identification of suitable auxiliary data available over the entire study area. From 2003 to 2013, the topographic survey institution of RLP conducted an airborne laserscanning acquisition over the entire federal state during leaf-off conditions in order to derive a countrywide digital terrain model (DTM) as well as a digital surface model (DSM). For this study, the recorded ALS data was used to create a canopy height model (CHM) in raster format, providing discrete information about the canopy surface height of the forest area in a spatial resolution of 5 meters (Fig. 2, *top*). The CHM was calculated as the difference between the digital terrain model and the digital surface model that were derived by a Delauney interpolation of the ground and first ALS pulses respectively. A more detailed description of the procedure can be found in Hill *et al.* [34]. The CHM provided the most valuable information to be used in the OLS regression model for predicting the timber volume on the plot and cluster level. However, it should be noted that the prolonged acquisition period of the ALS campaign led to the possibility of poor temporal alignment with the BWI3 survey, sometimes up to 10 years. In addition, the quality of the CHM varied substantially as ALS technology evolved over the years. For example, the ALS acquisitions recorded in 2002 and 2003 exhibited particularly poor quality with about only 0.04 point per m^2, whereas more recent datasets contained more than 5 points per m^2. Furthermore, CHM information was not available at 16 sample locations due to sensor failures. These plots were deleted from the sampling frame and treated as missing at random. This assumption was considered to be reasonable as the respective sample locations did not exclude specific forest structures.

5.4.2. Tree species map

Additional auxiliary data was derived from a countrywide satellite-based classification map predicting the five main tree species [35], i.e. European beech, Sessile and Pedunculate oak, Norway spruce, Douglas fir and Scots pine (Fig. 2, *bottom*). The tree species map has a grid size of 5x5 m and was calculated from 22 bi-temporal satellite images (SPOT5 and RapidEye) using a spatially adaptive classification algorithm [36]. As timber volume estimation on the tree level is often based on species-specific biomass and volume equations, the use of tree species information has often been stated as a key factor for improving the precision of timber volume estimates [37]. In this respect, incorporating the tree species map was particularly attractive as it predicts five of the seven tree species that are used in the BWI3 taper functions [27] to calculate the timber volume of a sample tree. However, due to unavailable satellite data, the tree species map excluded one large patch with an area of 415 km^2 in the south-west part of RLP covering an entire FA unit consisting of 10 FR units. In 9 additional FR units, the tree species information was also missing for a subset of the sample locations due to two additional patches with areas of 76 km^2 and 100 km^2 respectively in the northern part of RLP. For these 19 FR units, small area estimation was thus restricted to using only the available CHM information in the regression model. Thus, 411 of 5791 sample locations (approximately 7%) used to fit the regression model were affected by missing tree species information. A summary of the sample sizes and missing auxiliary data for both the CHM and the tree species map is provided in Table 2.

Table 2. Sample size for each phase in entire study area. $n_{\{1,2\},plots}$: number of plots. $n_{\{1,2\}}$: number of clusters. TSPEC: tree species map information.

Sampling frame	$n_{1,plot}$	n_1	$n_{2,plot}$	n_2
communal and state forest	96'854	33'365	5791	2055
missing CHM	18	10	0	0
missing TSPEC	7060	3587	414	385
missing CHM *and* TSPEC	3	2	0	0
missing CHM *or* TSPEC	7075	3595	414	385

5.5. Calculation of the explanatory variables

5.5.1. Canopy height model

The continuous explanatory variables derived from the CHM were the mean canopy height (*meanheight*) and the standard deviation (*stddev*). The quantities were calculated by evaluating the raster values around each sample location within a circle with a predefined radius of 12 meters, i.e. the support. In order to correct for edge effects at the forest border, the intersection of each support area to the state and communal forest area was determined using a polygon mask provided by the state forest service. The percentage of the support within the forest layer was used as the weight $w(x_l)$ introduced in Eq. 10 in order to derive the weighted mean of the explanatory variables on the cluster level. Neglecting the support adjustment would deteriorate the coherence between explanatory variables computed at the forest boundary and the corresponding local density that already includes a potential boundary adjustment, thus introducing unnecessary noise to the model. The boundary adjustment to the support also makes the sampling frame more consistent for the different data sources (Section 5.3).

The ALS acquisition year (*ALSyear*) was added as a categorical variable in order to account for the time lag with the terrestrial survey as well as to help explain the heterogeneity in the data introduced by the varying ALS quality. In 2008, a sensor error produced particularly poor ALS quality so the year was divided accordingly into two factor levels, denoted *2008_1* and *2008*. Furthermore, in order to increase the number of observations per factor level the years 2006 and 2007 were pooled together and the same was done for 2012 and 2013. The result was nine factor levels denoted as *2002*, *2003*, *2007*, *2008_1*, *2008*, *2009*, *2010*, *2011* and *2012*.

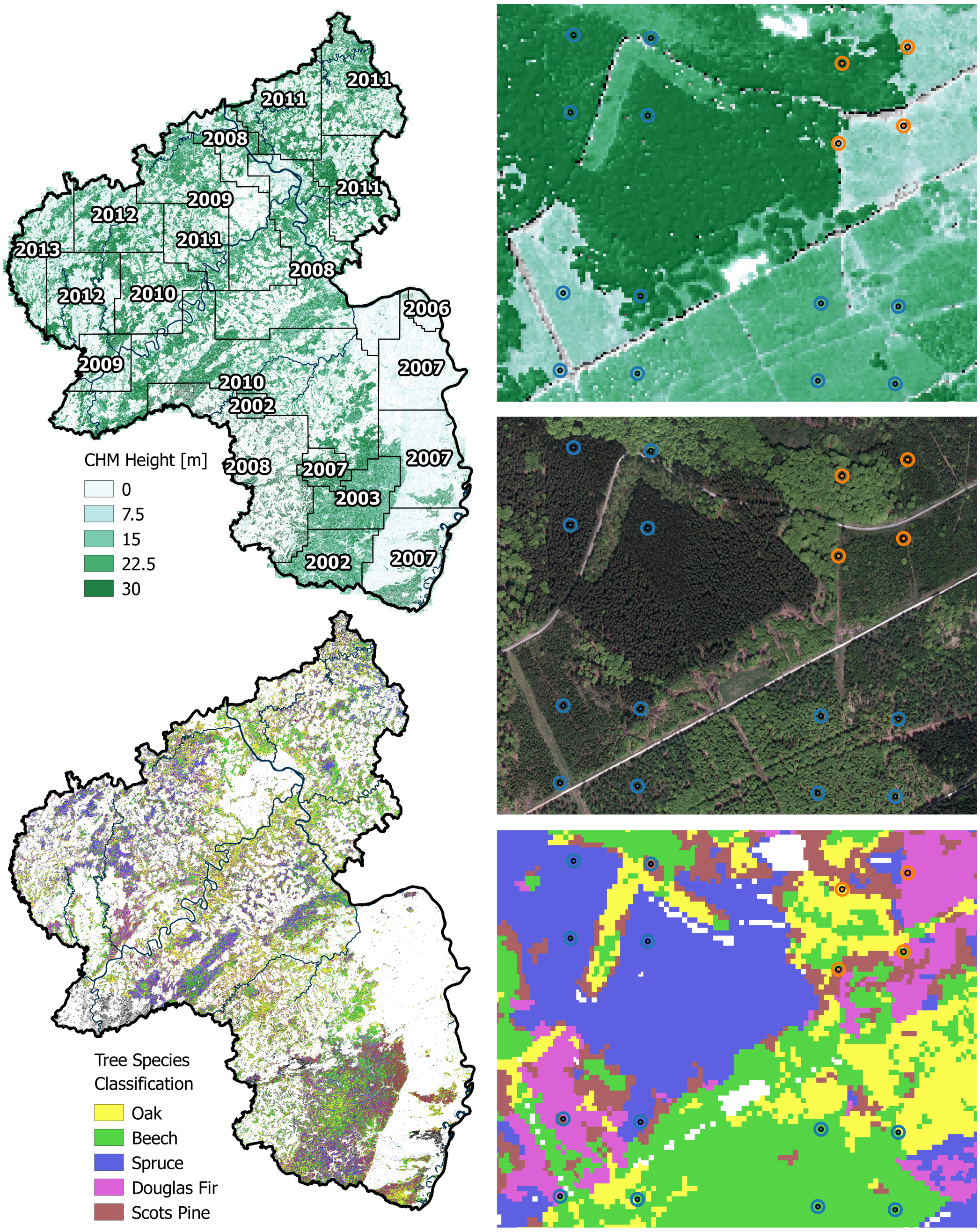

Figure 2. Left: CHM (top) and tree species classification map (bottom) available on the federal state level. Right: Magnified illustration of the supports used to derive the explanatory variables from the auxiliary data.

5.5.2. Tree species map

The tree species map was used to predict the main tree species at each sample plot which served as an additional categorical variable called *treespecies*. This involved two consecutive processing steps. In the first step, one of the five tree species was assigned to a sample location if 100% of the raster values within the edge-corrected support were classified as that species. Otherwise, the sample location was assigned the value 'mixed'. Likewise for the CHM variables, the support radius was 12 meters although the use of different support sizes for each explanatory variable would be in agreement with the two-phase estimators presented in Section 4.2. When using the *treespecies* variable in a regression model, the support size and the percentage threshold parameters had to be optimized in order to minimize the variance within each level which subsequently leads to improved model precision. A detailed analysis and description of the optimal parameter processing for the explanatory variables of the present data set is provided in Hill *et al.* [34]. In a second step, the *treespecies* variable was also passed through a calibration model in order to reduce the effects of misclassification errors on the regression model coefficients and to increase model accuracy. The calibration model consisted of a decision tree from a random forest algorithm [38] that was trained to predict the actual main plot tree species (known for all terrestrial plots) based on available auxiliary variables. These variables were the predicted *treespecies* variable, the mean canopy height and standard deviation of the CHM, as well as the proportion of coniferous trees estimated from the classification map and the growing region derived from a polygon map. The algorithm was grown with 2000 trees considering 3 of the predictors for each split. We thus applied this calibration model to the *treespecies* variable derived at all sample locations s_1. Table 3 gives the classification accuracies [39] of the *treespecies* variable after calibration.

Table 3. Classification accuracies of the *treespecies* variable before and after calibration. n_{ref}: number of terrestrial reference plots. n_{class}: number of classified plots.

Main plot species	Producer's accuracy[%]	User's accuracy[%]	n_{ref}	n_{class}
Beech	22.31	47.02	883	419
Douglas Fir	24.78	48.72	230	117
Oak	11.07	48.48	289	66
Spruce	53.15	61.13	651	566
Scots Pine	22.91	46.07	179	89
Mixed	84.49	64.53	3152	4127
	Overall accuracy: 61.96%		5384	5384

5.6. Regression Model

The model selection process for this study required a substantial time commitment due to sophisticated challenges such as: a) the heterogeneity of the remote sensing data, b) the identification of the optimal support sizes under angle count sampling, and c) the incorporation of tree species information. Here, only a summary of the extensive analysis that was performed is provided but the reader can refer to Hill *et al.* [34] if more details are desired.

The model with highest adjusted R^2 and lowest RMSE was achieved using *meanheight*, *meanheight*^{2}, *stddev*, *ALSyear* and *treespecies* as main effects, and including interaction terms between *meanheight* and *ALSyear*, *stddev* and *ALSyear*, *meanheight* and *stddev*, and *meanheight* and *treespecies*. Summary information about the adjusted R^2, RMSE and RMSE% of the selected models is provided in Table 4. As the two-phase estimators described in Section 4.2 derive and apply the regression coefficients and the residuals on the aggregated cluster level, we re-evaluated the model as used in the estimators on the cluster level (formulas given in Appendix) and found improved model fits compared to the plot level (adjusted R^2 of 0.59 and RMSE of 101.61 m^3/ha and 33.6%). The stratification by the ALS acquisition year substantially improved the model fit, indicating that it is an effective means in accounting for the noise in the data caused by ALS quality variations and time-gaps between the ALS and the terrestrial survey. However, the stratification led to a highly unbalanced data set when a further *treespecies* stratification was included. For this reason, a individual species modeling within each *ALSyear* stratum remained infeasible, but might have further improved the model fit. An additional evaluation of the model's performance within each ALS acquisition year stratum revealed that the quality of the model fit substantially varied between the strata (Table 5). In particular, values above the overall adjusted R^2 were higher in ALS acquisition years close to the terrestrial survey date compared to years with larger time gaps.

As described in Section 5.4.2, the information of the tree species classification map was missing within 1 FA and 19 FR units. For these small area units, we applied the regression model without the *treespecies* variable (Table 4, reduced model). However, the adjusted R^2s of the full and reduced model were found to be very similar on both the plot and cluster level. This implied that the variance reduction of the reduced model when applied to the two-phase estimators would likely be comparable to that of the full model, which is why a joint evaluation of the estimation results was performed (Section 6).

Table 4. Model fit specifications for the two OLS regression models on the cluster level. Interaction terms are indicated by ':'. () give the respective values on the plot level.

model terms	model	R^2_{adj}	RMSE	RMSE%
meanheight + stddev + meanheight^2 + treespecies + ALSyear + meanheight:treespecies + meanheight:ALSyear + meanheight:stddev + stddev:ALSyear	full model	0.58 (0.48)	90.11 (139.22)	29.76 (45.98)
meanheight + stddev + meanheight^2 + ALSyear + meanheight:ALSyear + meanheight:stddev + stddev:ALSyear	reduced model	0.55 (0.45)	95.23 (144.13)	31.65 (47.60)

Table 5. R^2, RMSE and RMSE% on the cluster level of the full regression model within ALS acquisition year strata (*ALSyear*). $Area_{ALSyear}$: Area covered by ALS acquisition given in km^2. n: sample size of validation data. () give the respective values on the plot level.

ALSyear	$Area_{ALSyear}$	R^2	RMSE	RMSE%	n
2012	2807	0.65 (0.61)	98.52 (135.84)	29.62 (44.87)	156 (408)
2011	4361	0.60 (0.57)	96.89 (146.21)	29.66 (48.29)	354 (883)
2010	4182	0.64 (0.51)	76.38 (120.90)	27.57 (39.93)	420 (1171)
2009	2100	0.53 (0.42)	92.22 (133.42)	33.31 (44.07)	218 (559)
2008	2968	0.61 (0.48)	87.10 (130.38)	32.20 (43.06)	247 (701)
2008_1	2116	0.43 (0.33)	117.99 (175.43)	33.64 (57.94)	157 (394)
2007	3498	0.56 (0.46)	82.43 (136.47)	26.57 (45.08)	135 (418)
2003	602	0.34 (0.27)	85.92 (154.48)	27.31 (51.02)	145 (529)
2002	775	0.52 (0.44)	87.25 (141.55)	27.22 (46.75)	97 (314)

Concerning the existence of outliers or leverage points in the training set for the model, it should be noted that it is more problematic for PSMALL, PSYNTH and EXTPSYNTH to simply remove them as one might be inclined to do in a model-dependent context. Strictly speaking, outlier removal in the design-based context essentially means that those plots, and implicitly any potentially similar plots that were not realized in the selected sample, have been removed from the sampling frame and are no longer considered part of the forest area of interest. While this may be valid for some obvious typos or measurement errors, it is generally not advisable to manipulate the sampling frame after observing data collected from it, especially when the observation in question lies within the small area of interest. However, for sake of completeness, we conducted an analysis of influential observations

[40, pp. 160–167] on the plot level for the full regression model. We calculated the leverage values and found that 10% of all observations exceeding a predefined critical threshold, i.e. twice the average of the hat matrix diagonal entries. Further investigation revealed that several leverage points showed unusually large *meanheight* values compared to their respective timber volume densities. They tended to occur in ALS acquisition years with longer time gaps to the terrestrial survey date and were thus more likely caused by harvesting activities in the sample plot area. Although these areas likely affected by harvest should clearly not be removed from the sampling frame, it does provided more justification for the inclusion of the *ALSyear* variable to mitigate these effects.

6. Results

6.1. General estimation results

An application of the SRS, PSMALL and EXTPSYNTH estimator was not feasible for 17 of all 405 FR-units due to an insufficient terrestrial sample size of $n_{2,G} < 2$. We further restricted the calculation of the PSMALL and EXTPSYNTH estimator to small area units with a minimum terrestrial sample size of $n_{2,G} \geq 4$ to avoid unstable estimates. This affected 65 additional FR units and limited unbiased two-phase estimations to 321 (79%) of the 405 FR units. It should be noted that also the PSYNTH estimator could not be applied for 2 FR-units since $n_{1,G} < 2$. Due to substantially larger sample sizes, all estimators could however be applied to all 45 FA units. The average value and the range of the mean timber volume estimates over the evaluated FA and FR units turned out to be very similar between all estimators (Table 6). An additional pairwise comparison of the 95% confidence intervals revealed that the four estimators did in fact not produce statistically different point estimates for all FA and FR units. This confirmed that the differences between the estimators are solely found in the precision which they provide for the point estimates.

Table 6. Descriptive summary of point estimates and estimation errors on the two forest district levels. N_u: number of evaluated small area units.

District level	Estimator		Point estimates			error$_{[\%]}$		
			mean	min	max	mean	min	max
FA	SRS	(N_u=45)	300.16	215.91	392.84	6.69	3.87	13.21
	PSMALL	(N_u=45)	307.29	209.26	417.10	5.16	3.46	14.33
	EXTPSYNTH	(N_u=45)	307.27	209.01	415.02	4.78	3.25	13.88
	PSYNTH	(N_u=45)	306.90	223.51	409.92	2.34	1.54	3.95
FR	SRS	(N_u=388)	301.83	99.89	612.13	18.32	0.34	104.97
	PSMALL	(N_u=321)	308.15	159.64	568.67	12.24	3.48	44.94
	EXTPSYNTH	(N_u=321)	308.38	154.07	544.34	11.34	3.60	40.91
	PSYNTH	(N_u=403)	307.82	166.01	444.29	4.65	2.56	62.51

6.2. Estimation error

On both small area levels, the design-unbiased estimators PSMALL and EXTPSYNTH led to a substantial reduction in the estimation error compared to the SRS estimator (Fig. 3). On the FA level, the SRS estimator yielded an estimation error of 6.7% on average compared to 5.2% and 4.8% under EXTPSYNTH and PSMALL respectively (Table 6). The cumulative error distribution (Fig. 3, left) reveals that under the SRS estimator, errors less than 5% were achieved for 17% of the FA units (8 of 45). This proportion could be increased to 62% (28 FA units) and 73% (33 FA units) by application of the PSMALL and EXTPSYNTH estimator. 95% of all estimates exhibited errors less than 9.5% under the SRS estimator and less than 6.6% when using PSMALL or EXTPSYNTH. Estimation errors higher than 10% only appeared twice for each of the three estimators.

Although the estimation errors were substantially larger overall on the FR level compared to the FA level due to smaller sample sizes, the error reduction from SRS by PSMALL and EXTPSYNTH were even more pronounced (Fig. 3, right). The average error under the SRS estimator was 18.3%, while it was 11.3% and 12.2% under PSMALL and EXTPSYNTH (Table 6). Errors smaller than 10% were achieved for 15% of the FR units by the SRS estimator, and for 46% by the PSMALL and PSYNTH estimator. 95% of the 321 FR units where PSMALL and EXTPSYNTH could be applied exhibited errors less than 20%. In comparion, the SRS estimates resulted in errors less than 36.6% for 95% of the 388 FR units.

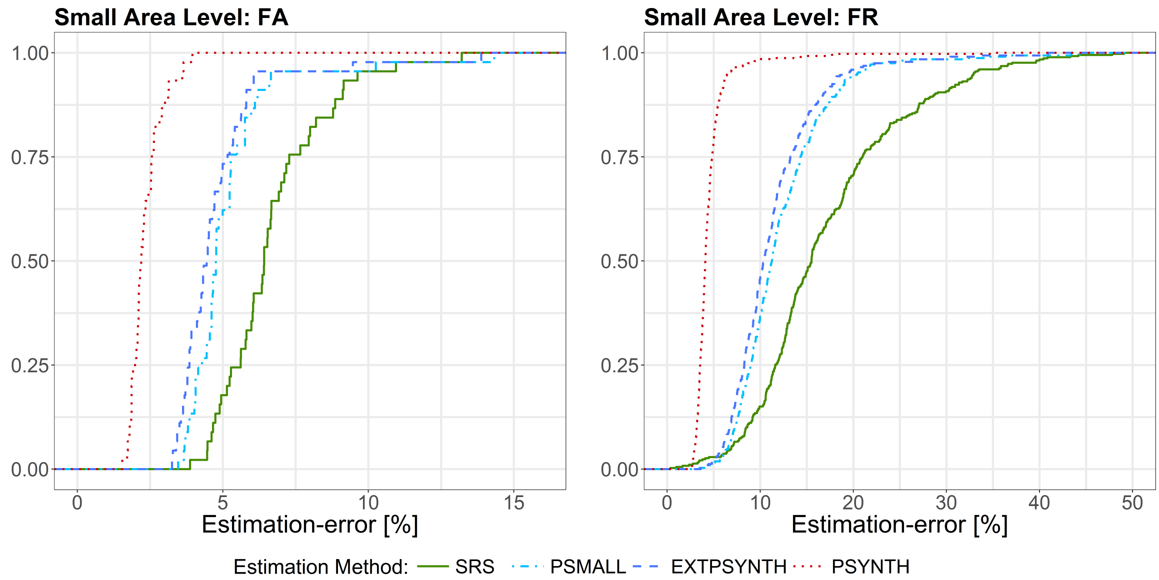

Figure 3. Cumulative distribution of estimation errors under SRS, PSMALL, EXTPSYNTH and the PSYNTH estimator. *Left*: Results for the 45 FA units. *Right*: Results for the 388 (SRS), 321 (PSMALL, EXTPSYNTH) and 403 (PSYNTH) FR units.

On both small area levels, the PSYNTH estimator resulted in much smaller estimation errors compared to PSMALL and EXTPSYNTH. This was as expected, since the PSYNTH variance estimate does not take the residual variation in each small area unit into account (Section 4.2.2). Compared to the asymptotically design-unbiased estimators PSMALL and EXTPSYNTH, the estimation errors produced by PSYNTH thus seem to be too optimistic. One should also recall that the estimates of the PSYNTH estimator are potentially design-biased.

6.3. Comparison of PSMALL and EXTPSYNTH

Figure 3 reveals that the error distribution of PSMALL and EXTPSYNTH are very similar, with PSMALL showing marginally higher estimation errors. In order to investigate the differences between PSMALL and EXTPSYNTH, we compared the g-weight variances of both estimators for all 321 FR units (Fig. 4, left). As obvious, PSMALL yielded slightly larger variances for the vast majority of the estimates. As addressed in Section 4.2.3, one possible explanation for differences was the effect of one or more clusters not entirely being included in a small area unit, as this would constitute an assumption violation of the EXTPSYNTH estimator. This violation was actually observed in 155 of the 321 FR units (48%). We compared the variances of PSMALL and EXTPSYNTH for all small areas that did not have the violations using a Wilcoxon Signed-Rank Test [41] on a 5% significance level. This test was also performed pairwise for groups $n_{2,G} \leq 6$, $n_{2,G} > 6$ and $n_{2,G} > 10$. The distribution of variances from EXTPSYNTH was found to be highly significantly lower than that of PSMALL except for the group of $n_{2,G} > 10$. The latter was expected since the variances of both estimators are asymptotically equivalent under large terrestrial sample sizes $n_{2,G}$ within the small area [29, pp.17–18].

This was also confirmed by a visual comparison of the absolute differences in the variances (Fig. 4, right) which decreased with increasing terrestrial sample size. Performing the same comparison for small areas with violations also revealed the EXTPSYNTH variances to be significantly smaller than the respective PSMALL variances until sample sizes $n_{2,G} > 10$. Based on these investigations, it was not possible to determine whether the differences for sample sizes smaller than 10 were caused by the violations or just reflect the general tendency of EXTPSYNTH to produce smaller variances than PSMALL under small sample sizes. However, a visual inspection provided some evidence that the violations created a statistically significant influence on the EXTPSYNTH variance (Fig. 4, left, red diamonds) that makes it appear to be slightly over-optimistic. For sample sizes of $n_{2,G} < 6$, a weakly significant difference between the EXTPSYNTH variances of those small areas with violations and the EXTPSYNTH variances without violation was also indicated by an unpaired Wilcoxon Rank-Sum Test. However, the differences were still marginal and a comparison of the confidence intervals of PSMALL and EXTPSYNTH revealed that the variance differences did not lead to statistically significant point estimates.

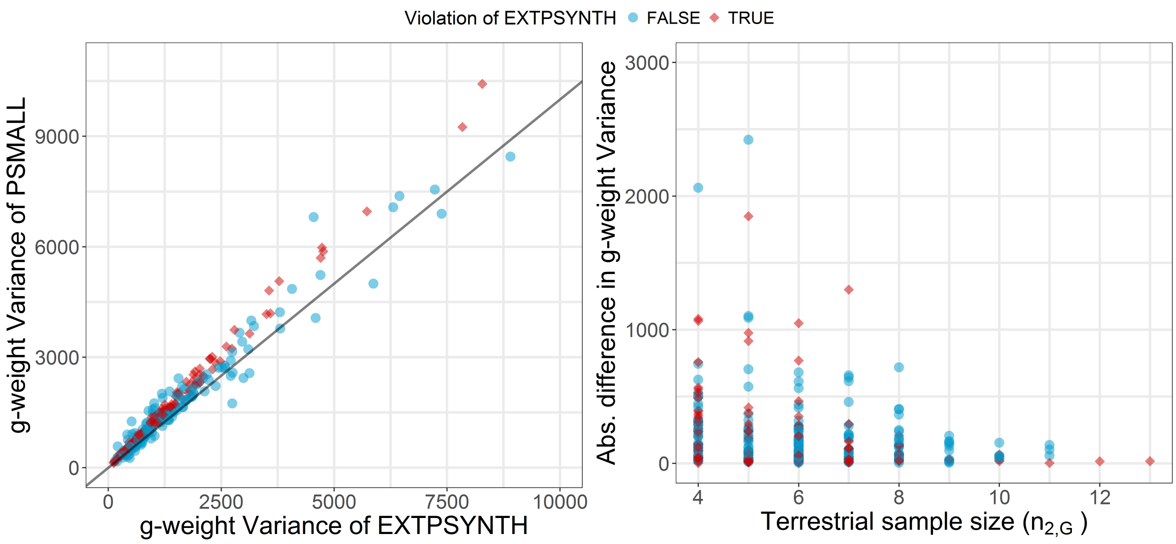

Figure 4. *Left*: Comparison of the g-weight variance between the PSMALL and the EXTPSYNTH estimator for the 321 FR units. *Right*: Difference in g-weight variance between the PSMALL and the EXTPSYNTH estimator in dependence of the terrestrial data ($n_{2,G}$) in the FR unit.

6.4. Variance reduction compared to SRS

The variance reduction relative to SRS for PSMALL and EXTPSYNTH are described in Figure 5 and Table 7. A direct comparison of the variances within the small area units revealed that the application of the design-unbiased estimators (PSMALL, EXTPSYNTH) led to a variance reduction compared to SRS in all FA units. In 75% of the FA units, the EXTPSYNTH estimator was able to reduce the variance by up to 54.1%. The reduction in variance can also be expressed in the relative efficiency values, which were 2.02 on average and ranged between 1.18 and 4.13 on the FA level. On FR level, the reduction in variance even reached values of 90% and relative efficiencies of 30 (Table 7 and Fig. 5). The PSMALL estimator again yielded slightly lower variance reductions and relative efficiencies due to the generally smaller variances of the EXTPSYNTH estimator (Section 6.3).

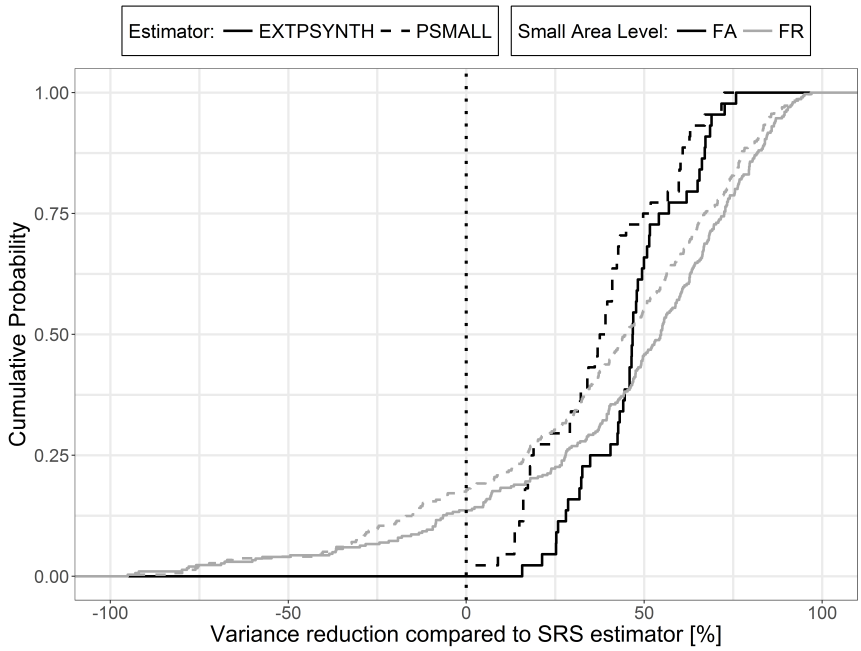

Figure 5. Cumulative distribution of variance reduction by the PSMALL and EXTPSYNTH compared to the SRS estimator for the 45 FA and 321 FR units.

Table 7. Descriptive summary of variance reduction compared to SRS and relative efficiencies on the two forest district levels. N_u: number of evaluated small area units.

District level	Estimator		Variance reduction [%]			relative efficiency		
			mean	min	max	mean	min	max
FA	PSMALL	(N_u=45)	33.51	2.6	72.5	1.74	1.03	3.64
	EXTPSYNTH	(N_u=45)	43.30	15.7	75.8	2.03	1.18	4.13
FR	PSMALL	(N_u=321)	12.48	-1203.9	96.8	2.54	0.08	31.61
	EXTPSYNTH	(N_u=321)	24.75	-892.7	97.0	2.95	0.10	33.70

Cases also occurred on the FR level where one or both two-phase estimators produced larger variance values than under the SRS estimator. This happened in 19% of the FR units under the EXTPSYNTH, and in 24% of the FR units under the PSMALL estimator. One possible reason for this was supposed to be a large residual variance due to a poor performance of the regression model within the small area unit. In order to investigate this hypothesis, we analyzed the three variance terms of the PSMALL estimator (eq. 14b), i.e. the variance introduced by the uncertainty of the regression coefficients (term 1), the variance caused by estimating the auxiliary means (term 2), and the variance of the model residuals (term 3). In general, the residual term is expected to make the largest contribution to the overall variance since it's sample size is based on $n_{2,G}$ whereas the auxiliary term and the coefficient term are based on larger sample sizes, i.e. $n_{1,G}$ and n_2 respectively. Figure 6 illustrates the share of the overall variance by the residual term of the PSMALL estimator scaled by the overall percentage reduction or increase of the variance compared to SRS for various small area sample sizes $n_{2,G}$. The residual term generally constitutes the dominating part of the PSMALL variance (around 84% on average). Although high residual term dominance does not necessary indicate that the PSMALL variance will be disproportionately large, as apparent from Figure 6 (*right*), the vast majority of the small areas where the PSMALL variance was larger than the SRS variance had residual terms contributing over 75% to the overall PSMALL variance. Furthermore, the magnitudes of the worst cases tended to occur in lower sample sizes. For example, of the FR units that saw variance increases where $n_{2,G} = 4$, the average increase was 272%, compared to 62% for FR units with $n_{2,G} > 4$ (Fig. 6, *left*). In comparison, the magnitude of the variance decreases were far more homogeneous than for the variance increases regardless of terrestrial sample size. Since $n_{2,G}$ is the same for PSMALL and SRS, this implies that the sum of square residuals for the model are likely larger than the sum of square local densities for the clusters in $s_{2,G}$ indicating the presence of outliers with large residuals in the problematic small areas. This situation is likely to arise when there was forest loss after the ALS scanning but before the terrestrial survey year.

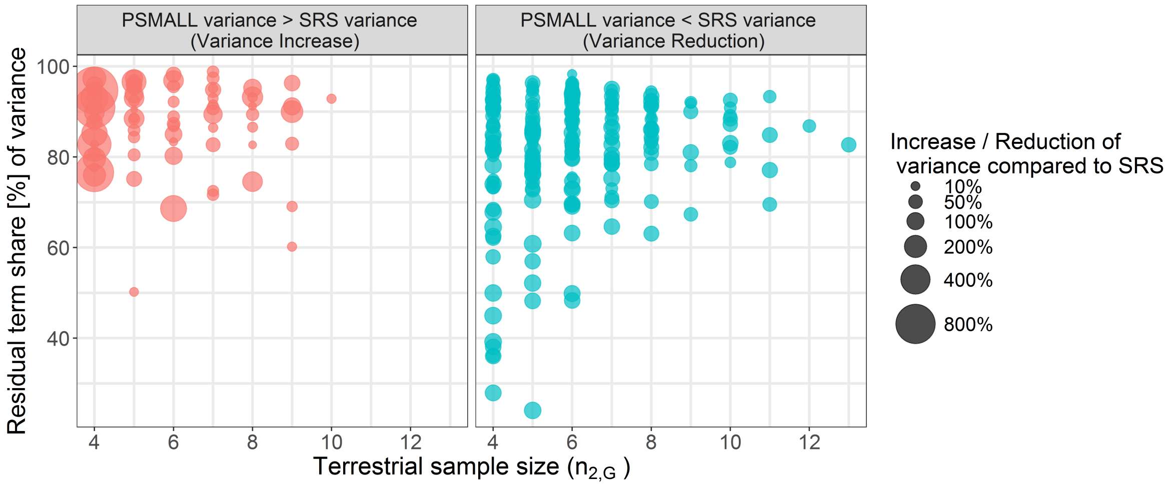

Figure 6. Share of the overall variance by the residual term of the PSMALL estimator for various small area sample sizes. Points are scaled by the overall percentage reduction/increase of the variance compared to SRS.

7. Discussion

7.1. Performance of estimators

The aim of this study was to investigate the performance of model-assisted design-based estimators for small area estimation of mean standing timber volume on two spatial forest management levels in Germany. It was of particular interest to gather information about the estimation error levels that can be attained using German NFI data that is characterized by low sampling intensities in the area of interests. To address these research questions, we applied the SRS, the PSMALL and the EXTPSYNTH estimators for cluster sampling to two forest management levels consisting of 45 and 405 small area units respectively in the German federal state of Rhineland-Palatinate.

Our study showed that on both small area levels, the PSMALL and the EXTPSYNTH estimators generally led to a substantial reduction in estimation error compared to the standard one-phase SRS estimator. On the upper management level (FA districts), PSMALL and EXTPSYNTH produced estimation errors smaller than 5% for 73% of the small areas compared to only 17% under the one-phase SRS estimator. The same level of precision could not be achieved on the lower management level (FR districts) primarily due to substantially smaller terrestrial sample sizes. However, in 95% of the FR units, the estimation errors could be limited to 20% compared to 40% under SRS. A pairwise comparison of the confidence intervals revealed that the estimators did not produce significantly different point estimates. The much smaller estimation errors of the PSYNTH estimator reflected the fact that it does not try to correct for potential bias in the point estimate which can lead to overly optimistic estimation errors and confidence intervals. One should thus prefer the unbiased estimates of PSMALL or EXTPSYNTH.

For several FR units, it was observed that the PSMALL and the EXTPSYNTH estimator can occasionally produce larger variances than the SRS estimator. It is important to note that this is in perfect agreement with the theory of both two-phase estimators and can theoretically appear if the residual variance in the small area, which generally constitutes the dominating part of the two-phase variance, turns out to be much higher than the variance of the terrestrial data in the small area. The empirical findings of our study suggest that such cases can particularly occur if moderate or poor model fits within a small area are combined with small terrestrial sample sizes (≤ 5) in the small area. A closer look on these small areas thus might reveal the reason for the poor prediction performance and help to improve the model fit. Nonetheless, it should be kept in mind that small terrestrial sample sizes can also cause the SRS estimator to not reflect the actual variation of the local density within a small area. In this case, the two-phase variance estimate might be larger but more realistic. Whereas a visual analysis of aerial images, remote sensing data or stand maps might give some further evidence for or against this hypothesis, a definite proof is practically infeasible.

We were also able to empirically confirm that the EXTPSYNTH estimator generally produces slightly smaller variances and estimation errors than the PSMALL estimator. This is most probably caused by marginally smaller model residuals due to the intercept adjustment to the terrestrial data in the small area unit, which is primarily a means to ensure the zero mean residual property of the EXTPSYNTH estimator. However, our analysis indicated that the difference between the two estimators is negligible for sample sizes ≥ 10 due to their asymptotic equivalency. We further investigated a potential impact on the EXTPSYNTH variance caused by the assumption violation that one or more clusters are not entirely included in the small area unit and found a slight but statistically significant tendency to be over-optimistic for sample sizes smaller than 6. More empirical evidence must be gathered before generalizing this as a rule of thumb for the application of the EXTPSYNTH under cluster sampling. It thus seems recommendable to prefer the EXTPSYNTH to the PSMALL estimator if its assumptions are not violated since it yields slightly smaller variances under mathematically soundness. Even if the differences between both estimators were marginal and did not lead to significantly different point estimates, PSMALL can serve as a safe alternative if the EXTPSYNTH assumption is violated. Aside from this, calculating both PSMALL and EXTPSYNTH and subsequently compare their results is always recommended to reveal suspicious deviations.

7.2. Auxiliary data

The auxiliary data used in our study were derived from two remote sensing sources, i.e. an ALS canopy height model and a tree species classification map. Likewise in many similar studies, the ALS mean canopy height proved to be the explanatory variable with highest predictive power. However, the large time-gaps of up to 10 years between the ALS acquisition and the terrestrial survey date caused the substantial introduction of artificial noise in the data. Whereas a post-stratification to the ALS acquisition years was an effective means to counteract the implied residual inflation, several leverage points were unambiguously caused by the temporal asynchronicity. Undetectable forest loss during the gap between the ALS acquisition and the NFI was also likely a cause for high residual variance in some small areas compared to the terrestrial data variance, which subsequently led to higher variances than the SRS estimator. As opposed to the ALS data, the availability of a country-wide tree species classification map has yet been unique among all German federal states. Whereas the study of Hill *et al.* [34] already showed that the tree species information was able to improve the model fit, it has yet not been used to its full potential. One reason for this was the impossibility of modeling individual tree species within each ALS acquisition year, which would add further explanatory power. Another reason was the lack of available satellite data for classification in some parts of the country, which led to missing values in the inventory data and restricted 19 FR units to a simpler regression model. Promising steps with respect to more up-to-date canopy height information have already been made, as the topographic survey institution of RLP will from this year on provide a country-wide canopy height model derived from aerial imagery acquisitions. These campaigns will in the future be conducted in a two-year period and allow to derive canopy height information matching the dates of terrestrial forest inventories. A study of Kirchhoefer *et al.* [42] recently indicated that similar model performance for German NFI data can be achieved using such imagery-based canopy height models. Due to the improved coverage and repetition rate of the Sentinel-2 satellite [43], the tree species classification map will in the future be updated each year. We consider these alternative auxiliary data sources to also solve the problem of missing explanatory variables at inventory plots. One could also make use of the exhaustive information within the two-phase estimators by using the true the auxiliary means [19,24], which could further decrease estimation errors. Previous studies of Mandallaz *et al.* [19] however showed that given a reasonable large sample size of the first phase, the differences in the estimation error are usually small. With respect to the substantial improvements in the temporal synchronicity between auxiliary and terrestrial inventory data, we consider the demonstrated double-sampling approach also to be very efficient for change estimation [44].

8. Conclusion

The study led to two major conclusions: (1) the EXTPSYNTH and PSMALL estimator generally achieved substantially smaller estimation errors on the two investigated forest district levels compared to the SRS estimator. The demonstrated double-sampling procedure thus constitutes a major contribution to an increase in value of the existing German NFI data on the federal state level. However, it is not possible to conclude from our study results alone whether the realized error levels are already sufficient enough in order to support forest planning decisions. Thus, further investigations are necessary in close cooperation with the forest authorities. A first study will concentrate on testing the EXTPSYNTH and PSMALL confidence intervals as a validation source for the stand-wise inventories. (2) Despite the quality restrictions in the ALS data and the tree species map, the two data sources were found to be well suited to model the mean timber volume on plot and cluster levels. With respect to frequently updated aerial canopy height models and tree species maps, it will thus be of interest to investigate the model and estimation performances that can be expected for future applications. In this framework, the incorporation of additional auxiliary data and the extension to change estimation seem the reasonable next steps to be explored towards an operational implementation of the demonstrated double-sampling procedure.

Acknowledgments: We want to express our gratitude to Prof. H. Heinimann (Chair of Land Use Engineering, ETH Zurich) for supporting this study. We also want to explicitly thank Dr. Johannes Stoffels and Dr. Henning Buddenbaum from the Environmental Sensing and Geoinformatic Group of University of Trier for providing the ALS data and tree species classification map, and Dr. Kai Husmann and Dr. Christoph Fischer from the Northwest German Forest Research Institution Göttingen for their advice in processing the terrestrial inventory data. Special gratitude is also owed to Dr. Thomas Riedel from the Thünen Institute for providing the densified NFI sample grid, and Dr. Alexander Massey for proofreading.

Author Contributions: Andreas Hill conducted the study and wrote the manuscript. Daniel Mandallaz developed the design-based estimators and supported the statistical analysis. Joachim Langshausen supported the study on the part of the State Forest Service Rhineland-Palatinate and cross-checked the analysis and the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix

R-squared on cluster level

The R^2 on the cluster level is calculated using the number of plots $M(x)$ of each cluster in order to weight for the varying number of plots on which $Y_c(x)$ and $\hat{Y}_c(x)$ are based on.

$$R^2 = \frac{\sum_{x \in s_2} \left(\frac{M(x)}{\bar{M}_2}\right)^2 \left(\hat{Y}_c(x) - \hat{\bar{Y}}_c\right)^2}{\sum_{x \in s_2} \left(\frac{M(x)}{\bar{M}_2}\right)^2 \left(Y_c(x) - \hat{\bar{Y}}_c\right)^2}$$

$Y_c(x)$ and $\hat{Y}_c(x)$ are the predicted and observed local densities on the cluster level calculated according to Equations 2 and 12. $\hat{\hat{Y}}_c$ is the estimated sample mean corresponding to the weighted mean over all observed local densities on the cluster level (Eq. 8).

RMSE on cluster level

The same weights $M(x)$ are also applied to calculate the RMSE on the cluster level. n_2 is the number of clusters used in the modeling frame.

$$RMSE = \sqrt{\frac{1}{n_2} \sum_{x \in s_2} \left(\frac{M(x)}{\bar{M}_2} \right)^2 \left(\hat{Y}_c(x) - Y_c(x) \right)^2}$$

The *relative* or *normalized* RMSE is calculated by dividing the RMSE by the estimated sample mean $\hat{\hat{Y}}_c$:

$$RMSE[\%] = \frac{RMSE}{\hat{\hat{Y}}_c}$$

Note that the weights $\frac{M(x)}{\bar{M}_2} \equiv 1$ if the number of plots per cluster is constant.

References

1. Polley, H.; Schmitz, F.; Hennig, P.; Kroiher, F. Germany. In *National Forest Inventories - Pathways for Common Reporting*; Springer, 2010; chapter 13, pp. 223–243.
2. Thünen-Institut. Dritte Bundeswaldinventur 2012, 2014. Accessed: 2017-02-03.
3. Kuliešis, A.; Tomter, S.M.; Vidal, C.; Lanz, A. Estimates of stem wood increments in forest resources: comparison of different approaches in forest inventory: consequences for international reporting: case studyof European forests. *Annals of Forest Science* **2016**, *73*, 857–869.
4. Böckmann, T.; Saborowski, J.; Dahm, S.; Nagel, J.; Spellmann, H. A new conception for forest inventory in lower saxony. *Forst und Holz (Germany)* **1998**.
5. von Lüpke, N. Approaches for the Optimisation of Double Sampling for Stratification in Repeated Forest Inventories. PhD thesis, University of Göttingen, 2013.
6. Särndal, C.E.; Swensson, B.; Wretman, J. *Model assisted survey sampling*; Springer Science & Business Media, 2003.
7. Gregoire, T.G.; Valentine, H.T. *Sampling strategies for natural resources and the environment*; CRC Press, 2007.
8. Köhl, M.; Magnussen, S.S.; Marchetti, M. *Sampling methods, remote sensing and GIS multiresource forest inventory*; Springer Science & Business Media, 2006.
9. Mandallaz, D. *Sampling techniques for forest inventories*; CRC Press, 2008.
10. Saborowski, J.; Marx, A.; Nagel, J.; Böckmann, T. Double sampling for stratification in periodic inventories—Infinite population approach. *Forest ecology and management* **2010**, *260*, 1886–1895.
11. Grafström, A.; Schnell, S.; Saarela, S.; Hubbell, S.; Condit, R. The continuous population approach to forest inventories and use of information in the design. *Environmetrics* **2017**.
12. von Lüpke, N.; Hansen, J.; Saborowski, J. A Three-Phase Sampling Procedure for Continuous Forest Inventory with Partial Re-measurement and Updating of Terrestrial Sample Plots. *European Journal of Forest Research* **2012**, *131*, 1979–1990.
13. Massey, A.; Mandallaz, D.; Lanz, A. Integrating remote sensing and past inventory data under the new annual design of the Swiss National Forest Inventory using three-phase design-based regression estimation. *Canadian Journal of Forest Research* **2014**, *44*, 1177–1186.
14. Mandallaz, D. A three-phase sampling extension of the generalized regression estimator with partially exhaustive information. *Canadian Journal of Forest Research* **2013**, *44*, 383–388.
15. Rao, J.N. *Small-Area Estimation*; Wiley Online Library, 2015.
16. Breidenbach, J.; Astrup, R. Small area estimation of forest attributes in the Norwegian National Forest Inventory. *European Journal of Forest Research* **2012**, *131*, 1255–1267.

17. Goerndt, M.E.; Monleon, V.J.; Temesgen, H. A comparison of small-area estimation techniques to estimate selected stand attributes using LiDAR-derived auxiliary variables. *Canadian journal of forest research* **2011**, *41*, 1189–1201.
18. Steinmann, K.; Mandallaz, D.; Ginzler, C.; Lanz, A. Small area estimations of proportion of forest and timber volume combining Lidar data and stereo aerial images with terrestrial data. *Scandinavian journal of forest research* **2013**, *28*, 373–385.
19. Mandallaz, D.; Breschan, J.; Hill, A. 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. *Canadian Journal of Forest Research* **2013**, *43*, 1023–1031.
20. Koch, B. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing data for forest biomass assessment. *ISPRS Journal of Photogrammetry and Remote Sensing* **2010**, *65*, 581–590.
21. Naesset, E. Area-Based Inventory in Norway – From Innovation to an Operational Reality. In *Forest Applications of Airborne Laser Scanning - Concepts and Case Studies*; Springer, 2014; chapter 11, pp. 216–240.
22. Magnussen, S.; Mauro, F.; Breidenbach, J.; Lanz, A.; Kändler, G. Area-level analysis of forest inventory variables. *European Journal of Forest Research* **2017**, pp. 1–17.
23. Magnussen, S.; Mandallaz, D.; Breidenbach, J.; Lanz, A.; Ginzler, C. National forest inventories in the service of small area estimation of stem volume. *Canadian Journal of Forest Research* **2014**, *44*, 1079–1090.
24. Mandallaz, D. Design-based properties of some small-area estimators in forest inventory with two-phase sampling. *Canadian Journal of Forest Research* **2013**, *43*, 441–449.
25. Bundesministerium für Ernährung, L.u.V. Aufnahmeanweisung für die dritte Bundeswaldinventur BWI3 (2011 - 2012), 2011.
26. Bitterlich, W. *The relascope idea. Relative measurements in forestry.*; Commonwealth Agricultural Bureaux, 1984.
27. Kublin, E.; Breidenbach, J.; Kändler, G. A flexible stem taper and volume prediction method based on mixed-effects B-spline regression. *European journal of forest research* **2013**, *132*, 983–997.
28. Schmitz, F.; Polley, H.; Hennig, P.; Dunger, K.; Schwitzgebel, F. Die zweite Bundeswaldinventur - BWI2: Inventur- und Auswertmethoden, Bundesministerium fur Ernahrung, Land- Wirtschaft und Verbraucherschutz (Hrsg), 2008.
29. Mandallaz, D.; Hill, A.; Massey, A. Design-based properties of some small-area estimators in forest inventory with two-phase sampling - revised version. Technical report, Department of Environmental Systems Science, ETH Zurich, 2016.
30. Hill, A.; Massey, A. *forestinventory: Design-Based Global and Small-Area Estimations for Multiphase Forest Inventories. R package version 0.3.1*, 2017.
31. Gauer, J.; Aldinger, E. Waldökologische Naturräume Deutschlands-Wuchsgebiete. *Mitteilungen des Vereins für Forstliche Standortskunde und Forstpflanzenzüchtung* **2005**, *43*, 281–288.
32. LWaldG. *Landeswaldgesetz Rheinland-Pfalz (Forest Act Rhineland-Palatinate)*, 2000. Rhineland-Palatinate, Germany.
33. Lamprecht, S.; Hill, A.; Stoffels, J.; Udelhoven, T. A Machine Learning Method for Co-Registration and Individual Tree Matching of Forest Inventory and Airborne Laser Scanning Data. *Remote Sensing* **2017**, *9*.
34. Hill, A.; Buddenbaum, H.; Mandallaz, D. Combining canopy height and tree species map information for large scale timber volume estimations under strong heterogeneity of auxiliary data and variable sample plot sizes. *European Journal of Forest Research* **2018**, *in print*.
35. Stoffels, J.; Hill, J.; Sachtleber, T.; Mader, S.; Buddenbaum, H.; Stern, O.; Langshausen, J.; Dietz, J.; Ontrup, G. Satellite-Based Derivation of High-Resolution Forest Information Layers for Operational Forest Management. *Forests* **2015**, *6*, 1982–2013.
36. Stoffels, J.; Mader, S.; Hill, J.; Werner, W.; Ontrup, G. Satellite-based stand-wise forest cover type mapping using a spatially adaptive classification approach. *European journal of forest research* **2012**, *131*, 1071–1089.
37. White, J.C.; Coops, N.C.; Wulder, M.A.; Vastaranta, M.; Hilker, T.; Tompalski, P. Remote sensing technologies for enhancing forest inventories: A review. *Canadian Journal of Remote Sensing* **2016**, *42*, 619–641.
38. Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5–32.
39. Congalton, R.G.; Green, K. *Assessing the accuracy of remotely sensed data: principles and practices*; CRC press, 2008.

40. Fahrmeir, L.; Kneib, T.; Lang, S.; Marx, B. *Regression: models, methods and applications*; Springer Science & Business Media, 2013.
41. Wilcoxon, F.; Katti, S.; Wilcox, R.A. Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test. *Selected tables in mathematical statistics* **1970**, *1*, 171–259.
42. Kirchhoefer, M.; Schumacher, J.; Adler, P.; Kändler, G. Considerations towards a Novel Approach for Integrating Angle-Count Sampling Data in Remote Sensing Based Forest Inventories. *Forests* **2017**, *8*, 239.
43. ESA. Sentinel-2 earth observation mission, 2017. Accessed: 2017-03-29.
44. Massey, A.; Mandallaz, D. Design-based regression estimation of net change for forest inventories. *Canadian Journal of Forest Research* **2015**, *45*, 1775–1784, [https://doi.org/10.1139/cjfr-2015-0266].