

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 24th June 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 for future large-scale operational application by the example of timber volume estimation on two small-scale management levels (45 and 405 forest district units respectively) over the entire area of 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 one-phase simple random sampling 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 (UNFCCC) 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 forested area of Germany. In order to cover a large area of 114'191 km² [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 judgements from time-consuming standwise forest 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) carried out for forests owned by the state forest enterprise 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 [5,6]. One method which has proven to be efficient is double- or two-phase sampling [7–12]. 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 in various countries such as Canada [13], the USA [14], Switzerland [15], Italy [16] and Germany [17]. Recent studies from Grafström *et al.* [18] illustrated how to use the auxiliary information to determine optimised balanced terrestrial sample designs. Double-sampling has also been extended to triple-sampling estimation methods using auxiliary information derived at two different sampling intensities. An example can be found in von Lüpke *et al.* [5] 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.* [19] who developed a triple-sampling extension based on the ideas of Mandallaz [20] 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 [11] originally comes from official statistics. One such method that is commonly applied is known as indirect estimation [21], 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 [22–25], 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 [26,27]. Other studies have investigated area-level models, where the auxiliary information is only provided on the SA level [28]. 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 [22] 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.* [29] 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 objective of this study was to investigate whether the application of latest design-based small area estimation methods allow to use the German NFI data to produce estimates of acceptable precision

on two forest district levels. The methods were tested in the German federal state Rhineland-Palatinate. Three types of model-assisted design-based small area regression estimators were used to derive point and variance estimates of mean standing timber volume for 45 and 405 forest management units on the two respective district levels. The SA-estimators we considered were the *pseudo-small*, *extended pseudo-synthetic* and the *pseudo-synthetic* design-based small area estimator suggested by Mandallaz [30] and Mandallaz *et al.* [25]. 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 partially 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 about the potential of the suggested small area estimation procedure and the incorporated auxiliary information for future operational large scale application.

2. Terrestrial sampling design of the German NFI

The German NFI (Bundeswaldinventur, BWI) is a periodic inventory that is carried out every 10 years over the entire forest area of Germany. The third and 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 [31]. In the field survey of the BWI3, sample trees for timber volume estimation are selected according to the angle count sampling technique [32]. Angle count sampling is a visual selection procedure conducted via an optical instrument (*Relascope*). At a sample point x , a tree is selected as a sample tree if its apparent diameter at the height of 1.3 m (DBH) and the associated angle α_i (in radiant) appears larger than a limit angle α when viewed through the Relascope [12]. The limit angle α can be expressed by the so-called basal area factor (BAF), i.e. $BAF = 10^2 \sin^2(\frac{\alpha}{2})$. Consequently, each tree i is assigned with an individual radius $R_i = \frac{DBH_i}{2\sqrt{BAF}}$ which is the distance between the tree location and the sample plot center. A tree is selected if R_i is equal to or smaller than the so-called limiting distance R that results from chosen BAF (i.e., α). One characteristic of angle count sampling is that a sample plot does no longer has a fixed radius in which sample trees are selected, but each tree creates its individual radius R_i that is - likewise the associated inclusion probability - proportional to its basal area. For this reason, the angle count method is also referred to as *variable radius plot sampling* and implements the *probability proportional to size (PPS)* concept. The BWI3 uses a basal area factor 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 [31]. A further inventory threshold for a tree to be recorded is a 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 [33].

129 3. Double sampling in the infinite population approach

130 3.0.1. One- and Two-Phase Sampling in the Infinite Population Approach

131 The estimators used in this study have been proposed by [25,30] and derive their mathematical
 132 properties under the so-called infinite population approach. Therefore, we shall first provide a short
 133 introduction into this general estimation framework. We start by assuming that the population P of
 134 trees $i \in 1, 2, \dots, N$ within a forest of interest F is exactly defined, and each tree i has a response variable
 135 Y_i (e.g. its timber volume) that can be used to define the population mean Y (e.g., the average timber
 136 volume per unit area) over F . Since a full census of all tree population individuals is almost never
 137 feasible, Y has to be estimated based on a sample. In the infinite population approach this sample is a
 138 set of points or locations x distributed independently and uniformly over the set of all possible points
 139 in F . Each point x has an associated local density $Y(x)$ (e.g., the timber volume per unit area) whose
 140 spatial distribution is given by a fixed (i.e. non stochastic) piecewise constant function. The population
 141 mean Y is mathematically equivalent to the integral of the local density function surface divided by
 142 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
 143 corresponds to a spatial mean. Since the actual local density function is unobserved in its entirety, one
 144 estimates Y by taking a sample s_2 consisting of n_2 points and measuring each of their respective local
 145 densities. This sampling procedure is often referred to as *one-phase sampling* (OPS) and s_2 is referred
 146 to as the terrestrial inventory. In contrast to the one-phase approach, *two-phase* or *double-sampling*
 147 procedures use information from two nested samples (phases). Practically speaking, the terrestrial
 148 inventory s_2 is embedded in a large phase s_1 comprising n_1 sample locations that each provide a set of
 149 explanatory variables described by the column vector $Z(x) = (z(x)_1, z(x)_2, \dots, z(x)_p)^\top$ at each point
 150 $x \in s_1$. These explanatory variables are derived from auxiliary information that is available in high
 151 quantity within the forest F . For every $x \in s_1$, $Z(x)$ is transformed into a prediction $\hat{Y}(x)$ of $Y(x)$
 152 using the choice of some prediction model. The basic idea of this method is to boost the sample size
 153 by providing a large sample of less precise but cheaper predictions of $Y(x)$ in s_1 and to correct any
 154 possible model bias, i.e., $\mathbb{E}(Y(x) - \hat{Y}(x))$, using the subsample of terrestrial inventory units where the
 155 value of $Y(x)$ is observed. In this context, it is also important to note that the response and auxiliary
 156 variables are assumed to be error-free and the resulting errors for the point estimates reflect only the
 157 uncertainty due to sampling.

158 4. Estimators

159 4.1. Design-based one-phase estimator for cluster sampling (SRS)

160 The one-phase estimator for cluster sampling (SRS) constitutes the *status quo* that is currently
 161 applied under the existing one-phase sampling design of the German NFI in order to obtain point
 162 and variance estimates for the mean timber volume of a given estimation unit. In order to provide
 163 all estimators in the infinite population framework and ensure a consistent terminology with the
 164 two-phase estimators in Section 4.2, we will introduce the SRS estimator that is applied in the BWI3
 165 algorithms [34] in the form given in Mandallaz [12] and Mandallaz *et al.* [35].

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

$$171 M(x) = \sum_{l=1}^M I_F(x_l) \quad \text{where} \quad I_F(x_l) = \begin{cases} 1 & \text{if } x_l \in F \\ 0 & \text{if } x_l \notin F \end{cases} \quad (1)$$

¹⁷¹ The local density on cluster level $Y_c(x)$, which is in our case the timber volume per hectare, is
¹⁷² then defined as the average of the individual sample plot densities $Y(x_l)$:

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

¹⁷³ The local density $Y(x_l)$ on individual sample plot level was calculated according to the description
¹⁷⁴ in Mandallaz [12], which can be rewritten for angle-count sampling technique applied in the BWI3.
¹⁷⁵ The general form of $Y(x)$ in Mandallaz [12] is given as the Horwitz-Thompson estimator

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

¹⁷⁶ where Y_i is in our case the timber volume of the tree i recorded at sample location x in m^3 . Each tree
¹⁷⁷ has an inclusion probability π_i that is well defined as the proportion of its inclusion circle area $\lambda(K_i)$
¹⁷⁸ within the forest area $\lambda(F)$, i.e. via their geometric intersection:

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

¹⁷⁹ 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
¹⁸⁰ as *limiting distance*), where $cf_{i,corr}$ is the original counting factor cf corrected for potential boundary
¹⁸¹ effects at the forest border. In case of angle-count sampling, we can rewrite π_i as

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

¹⁸² 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
¹⁸³ the corrected counting factor:

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

¹⁸⁴ Eq. 5 in Eq. 3 yields the rewritten form of $Y(x_l)$ for angle count sampling that conforms to the
¹⁸⁵ definition used in the BWI3 algorithms [34]:

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

¹⁸⁶ where nha_i is the number of trees per hectare represented by tree i . The local densities on cluster level
¹⁸⁷ can then be used to derive the estimated spatial mean \hat{Y}_c and its estimated variance $\hat{\text{V}}(\hat{Y}_c)$ for any
¹⁸⁸ given spatial unit for which $n_2 \geq 2$ (n_2 denoting the number of clusters):

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

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

¹⁸⁹ with $\bar{M}_2 = \frac{\sum_{x \in s_2} M(x)}{n_2}$.

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

¹⁹¹ All three considered small area estimators use ordinary least square (OLS) regression models to
¹⁹² produce predictions of the local density $Y_c(x)$ directly on the cluster level c . We consider the internal
¹⁹³ 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:

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

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

$\mathbf{Z}_c(x)$ is the column 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, \dots, x_l (Eq. 10). The weight $w(x_l)$ is the proportion of the extraction area (support) within the forest F used to derive the explanatory variables from the raw auxiliary information.

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

The estimated design-based variance-covariance matrix $\hat{\Sigma}_{\hat{\beta}_{c,s_2}}$ 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

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

with

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

being the empirical model residuals at the cluster level. By construction of OLS, the empirical model residuals satisfy that they average to zero, i.e. $\frac{\sum_{x \in s_2} M(x) \hat{R}_c(x)}{\sum_{x \in s_2} M(x)} = 0$, if $\mathbf{Z}_c(x)$ contains the intercept term. 1. Because the model is fitted internally, i.e. using the terrestrial sample s_2 of the inventory area, this leads to the important *zero mean residual property* of the theoretical model residuals over the inventory domain F , i.e. $\int_F R_c(x) dx$. Mandallaz [30] and Mandallaz *et al.* [25] used this property particular for the case of small area estimation to derive better variance estimates based on the g-weight technique adapted from the works of Särndal *et al.* [9].

In the following, we will give a short description of each small area estimator and refer to Mandallaz *et al.* [25], Mandallaz [30], Mandallaz *et al.* [35] if the reader requires additional mathematical details or proofs. The estimators have also been implemented in the open-source software-package *forestinventory* [36] in the statistical software *R* [37] which was used to compute all estimates in this study.

218

219 4.2.1. Pseudo Small Area Estimator (PSMALL)

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

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

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

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

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

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

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

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

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

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

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

234 The estimated design-based variance of $\hat{Y}_{c,G,PSMALL}$ is given by Eq. 14b. Basically, the first
 235 term constitutes the variance introduced by the uncertainty in the regression coefficients, whereas
 236 the second term expresses the variance caused by estimating the exact auxiliary mean in G using a
 237 non-exhaustive sample $s_{1,G}$. The third term is the variance of the model residuals and thus accounts for
 238 the inaccuracies of the model predictions. Note that the first term can also be rewritten using g-weights
 239 [35, pg.14] which ensures some beneficial calibration of the auxiliary variables to the first-phase sample.
 240

241

242 4.2.2. Pseudo Synthetic Estimator (PSYNTH)

243 The PSYNTH estimator is commonly applied when no terrestrial sample is available within
 244 the small area G (i.e. $n_{2,G} = 0$). The point estimate (Eq. 16a) is thus only based on the predictions
 245 generated by applying the globally derived regression coefficients to the small area cluster means of
 246 the explanatory variables $\hat{\mathbf{Z}}_{c,G}$. Note that the bias correction term using the empirical residuals (Eq.
 247 14a) can no longer be applied. Conditioned on the realized sample, the PSYNTH estimator can thus
 248 potentially have an unobservable design-based bias.

249

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

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

250 The contribution to the variance by the model residuals in small area G can also no longer be
 251 considered (Eq. 16b). As a result, the synthetic estimator will usually have a smaller variance than
 252 estimators that consider the model residuals, but at the cost of a potential bias. Note that the PSYNTH
 253 estimations are still design-based, but one purely has to rely on the validity of the regression model
 254 within the small area as it is the case in the model-dependent framework.
 255

256 4.2.3. Extended Pseudo Synthetic Estimator (EXTPSYNTH)

257 The EXTPSYNTH estimator (Eq. 17) has been proposed by Mandallaz [30] as a transformed
 258 version of the PSMALL estimator that has the form of the PSYNTH estimator but remains
 259 asymptotically design unbiased. It has the advantage that the mean of the empirical model residuals
 260 of the OLS regression model for the entire area F and the small area G are by construction both
 261 zero at the same time, i.e. $\bar{R}_c = \bar{R}_{c,G} = 0$. This is realized by *extending* the auxiliary vector $\mathbf{Z}_c(x)$
 262 by the indicator variable $I_{c,G}$ which takes the value 1 if the entire cluster lies within the small area
 263 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
 264 becomes $\mathbf{Z}_c^\top(x) = (\mathbf{Z}_c^\top(x), I_{c,G}(x))$ and the new regression coefficient using $\mathbf{Z}_c(x)$ instead of $\mathbf{Z}_c(x)$
 265 in Eq. 9 is denoted as $\hat{\boldsymbol{\theta}}_{s_2}$. All remaining components are calculated by plugging in $\mathbf{Z}_c(x)$ in Eq. 13.
 266 A decomposition of $\hat{\boldsymbol{\theta}}_{s_2}$ reveals that the residual correction term is now included in the regression
 267 coefficient $\hat{\boldsymbol{\theta}}_{s_2}$ [35].

$$\hat{Y}_{c,G,EXTPSYNTH} = \hat{\mathbf{Z}}_{c,G}^\top \hat{\boldsymbol{\theta}}_{c,s_2} \quad (17a)$$

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

268 However, it is important to note that $\bar{R}_{c,G} = 0$ under the extended regression model only holds if
 269 the sample plots x_1, \dots, x_l of a cluster are *all* either inside or outside the small area, i.e. $M_G(x) \equiv M(x)$,
 270 and thus $I_{c,G}(x) = \frac{M_G(x)}{M(x)}$ can only take the values 1 or 0. Mandallaz *et al.* [35] assumed that the
 271 effects on the estimates should be negligible as the number of occasions where $M_G(x) < M(x)$ was
 272 considered to be small in practical implementations. It was thus a further objective of this study to
 273 investigate the actual number of occurrences as well as effects of this phenomenon by comparing the
 274 estimates of EXTPSYNTH to those of PSMALL.

275 4.3. Measures of estimation accuracy

276 The estimation precision was quantified by the estimation error, which is the ratio of the standard
 277 error and the point estimate (here \hat{Y} stands for the point estimate produced under the various
 278 estimators):

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

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

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

²⁸³ The confidence intervals for the PSMALL and EXTPSYNTH estimates are calculated as:

$$CI_{1-\alpha}(\hat{Y}_{c,G,EXTPSYNTH}) = \hat{Y}_{c,G,EXTPSYNTH} \pm t_{n_2,G-1,1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y}_{c,G,EXTPSYNTH})} \quad (20a)$$

$$CI_{1-\alpha}(\hat{Y}_{c,G,PSMALL}) = \hat{Y}_{c,G,PSMALL} \pm t_{n_2,G-1,1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y}_{c,G,PSMALL})} \quad (20b)$$

²⁸⁴ For the PSYNTH estimates, the confidence intervals are

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

²⁸⁵ with p being the number of parameters used in the regression model including the intercept term.

²⁸⁶ In order to address the potential benefits of the small area estimators compared with the SRS
²⁸⁷ approach, we calculated the *relative efficiency* (*RE*, Eq. 22) which can be interpreted as the relative
²⁸⁸ sample size under SRS needed to achieve the variance under the double-sampling (DS) estimators.
²⁸⁹

$$RE = \frac{\hat{V}(\hat{Y}_{SRS})}{\hat{V}(\hat{Y}_{DS})} \quad (22)$$

²⁹⁰ where \hat{Y} stands for the point estimate produced under the respective estimator.

²⁹¹ 5. Case study

²⁹² 5.1. Study area and small area units

²⁹³ The German federal state Rhineland-Palatinate (*RLP*) is located in the western part of Germany
²⁹⁴ and borders Luxembourg, France and Belgium. With 42.3% (appr. 8400 km²) of the entire state area
²⁹⁵ (19850 km²) covered by forest, RLP is one of the two states with the highest forest coverage among all
²⁹⁶ federal states of Germany [2]. The forests of RLP are further characterised by a pronounced diversity in
²⁹⁷ bioclimatic growing conditions that have strong influence on the local growth dynamics as well as tree
²⁹⁸ species composition [38] and are further characterised by large variety of forest structures ranging from
²⁹⁹ characteristic oak coppices (Moselle valley), pure spruce, beech and scots pine forests (i.a. Hunsrück
³⁰⁰ and Palatinate forest) up to mixed forests comprising variable proportions of oak, larch, spruce, Scots
³⁰¹ pine and beech. Around 82% of the forest area in RLP are mixed forest stands and 69% of the forest
³⁰² area exhibit a multi-layered vertical structure. The forest area of RLP are divided into 3 ownership
³⁰³ classes, i.e. state forest (27%), municipal forest (46%) and privately owned forest (27%). The forest
³⁰⁴ service of RLP has the legal mandate to sustainably manage the state and municipal forest area (73% of
³⁰⁵ the entire forest area), including forest planning, harvesting and the sale of wood [39]. For this reason,
³⁰⁶ the entire forest area has been spatially organised in 3 main hierarchical management units (Figure 1).
³⁰⁷ On the upper level, RLP has been divided into 45 forest districts (Forstämter, *FA*), which are further
³⁰⁸ divided into a total number of 405 sub-districts (Forstreviere, *FR*). The next level are the forest stands
³⁰⁹ (104'184 in total) for which expert judgements are conducted by SFIs in a 5 to 10 year period in order
³¹⁰ to set up management strategies for the upcoming 10 years. The *FAs* and *FRs* constituted the SA units
³¹¹ for which design-based small area estimations of the mean standing timber volume were calculated
³¹² by incorporating the available terrestrial inventory data of the BWI3 in the estimators described in
³¹³ Section 4. The average area of the SA units was 43'777 ha on the *FA*-level, and 4624 ha on the *FR* level.

314 5.2. Terrestrial sample

315 Rhineland-Palatinate is covered by a 2x2 km inventory grid of the German NFI. In the last
316 inventory (BWI3) conducted in the year 2011 and 2012, timber volume information was derived
317 for 2810 clusters (8092 plots) in the field survey. The local timber volume density on the plot and
318 cluster level for this sample was consequently calculated according to Section 4.1. In the framework
319 of this survey, the plot center coordinates were re-measured with the differential global satellite
320 navigation system (DGPS) technique. Knowledge about the exact plot positions were considered
321 crucial to provide optimal comparability between the terrestrial observations and the information
322 derived from the auxiliary information. A comparison of the DGPS coordinates with the so-far used
323 target coordinates revealed that 90% of all horizontal deviations lay in the range of 25 meters. A
324 detailed analysis of horizontal DGPS errors in RLP by Lamprecht *et al.* [40] indicated that 80% of the
325 plots should not exceed horizontal DGPS errors of 8 meters. For 162 plots, the DGPS coordinates were
326 replaced by their target coordinates due to missingness or implausible values. The terrestrial sample
327 size $n_{2,G}$ within the FA units was 46 clusters on average and ranged between 11 and 64. Within the FR
328 units, $n_{2,G}$ was considerably smaller with an average of 5 clusters and a range between 0 and 13.

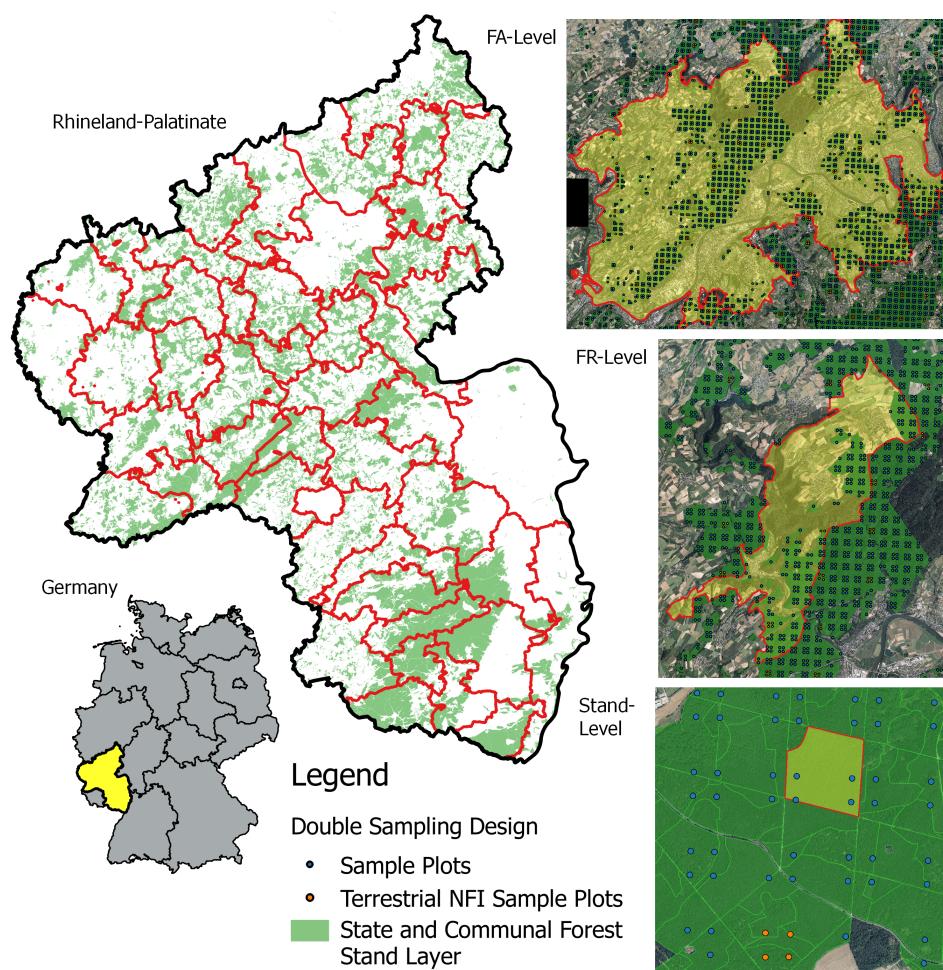


Figure 1. *Left:* Study area with delineated FA forest management units. *Right:* Example for each of the three management units (from top to bottom): FA, FR and forest stand unit overlayed with the extended double-sampling cluster design. Green: Forest stand polygon layer defining the state and municipal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

329 5.3. Extension to double-sampling design

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

Table 1. Descriptive statistics of the forest observed on NFI sample plots located within communal and state forest area ($n_2=5791$).

Variable	Mean	SD	Maximum
Timber Volume (m ³ /ha)	300.9	195.6	1375.3
Mean DBH (mm)	354.9	137.2	1123.2
Mean height (dm)	239.6	72.4	497.4
Mean stem density per hectare	101	114	1010

344 5.4. Auxiliary data

345 5.4.1. LiDAR canopy height model

346 A prerequisite for the application of the suggested two-phase small area estimators is the
347 identification of suitable auxiliary data available over the entire study area. From 2003 to 2013,
348 the topographic survey institution of RLP conducted an airborne laserscanning acquisition over the
349 entire federal state during leaf-off conditions in order to derive a countrywide digital terrain model
350 (DTM) as well as a digital surface model (DSM). For this study, the recorded ALS data was used to
351 create a canopy height model (CHM) in raster format, providing discrete information about the canopy
352 surface height of the forest area in a spatial resolution of 5 meters (Fig. 2, top). The CHM was calculated
353 as the difference between the digital terrain model and the digital surface model that were derived by
354 a Delauney interpolation of the ground and first ALS pulses respectively. A more detailed description
355 of the procedure can be found in Hill *et al.* [41]. The CHM provided the most valuable information to
356 be used in the OLS regression model for predicting the timber volume on the plot and cluster level.
357 However, it should be noted that the prolonged acquisition period of the ALS campaign led to the
358 possibility of poor temporal alignment with the BWI3 survey, sometimes up to 10 years. In addition,
359 the quality of the CHM varied substantially as ALS technology evolved over the years. For example,
360 the ALS acquisitions recorded in 2002 and 2003 exhibited particularly poor quality with about only 0.04
361 point per m², whereas more recent datasets contained more than 5 points per m². Furthermore, CHM
362 information was not available at 16 sample locations due to sensor failures. These plots were deleted
363 from the sampling frame and treated as missing at random. This assumption was considered to be
364 reasonable as the respective sample locations did not systematically exclude specific forest structures.

365 5.4.2. Tree species classification map

366 Additional auxiliary data was derived from a countrywide satellite-based classification map
367 predicting the five main tree species [42], i.e. European beech, Sessile and Pedunculate oak, Norway
368 spruce, Douglas fir and Scots pine (Fig. 2, bottom). The tree species map has a grid size of 5x5 m
369 and was calculated from 22 bi-temporal satellite images (SPOT5 and RapidEye) using a spatially
370 adaptive classification algorithm [43]. As timber volume estimation on the tree level is often based
371 on species-specific biomass and volume equations, the use of tree species information has often been
372 stated as a key factor for improving the precision of timber volume estimates [44]. In this respect,
373 incorporating the tree species map was particularly attractive as it predicts five of the seven tree species
374 that are used in the BWI3 taper functions [33] to calculate the timber volume of a sample tree. However,
375 due to unavailable satellite data, the tree species map excluded one large patch with an area of 415
376 km² in the south-west part of RLP covering an entire FA unit consisting of 10 FR units. In 9 additional
377 FR units, the tree species information was also missing for a subset of the sample locations due to two
378 additional patches with areas of 76 km² and 100 km² respectively in the northern part of RLP. For these
379 19 FR units, small area estimation was thus restricted to using only the available CHM information in
380 the regression model. Thus, 411 of 5791 sample locations (approximately 7%) used to fit the regression
381 model were affected by missing tree species information. A summary of the sample sizes and missing
382 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. $n_{\{1,2\},plot}$: 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
municipal 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

383 5.5. Calculation of the explanatory variables

384 5.5.1. Canopy height model

385 The continuous explanatory variables derived from the CHM were the mean canopy height
386 (*meanheight*) and the standard deviation (*stddev*). The quantities were calculated by evaluating the
387 raster values around each sample location within a circle with a predefined radius of 12 meters, i.e.
388 the support. In order to correct for edge effects at the forest border, the intersection of each support
389 area to the state and municipal forest area was determined using a polygon mask provided by the
390 state forest service. The percentage of the support within the forest layer was used as the weight
391 $w(x_l)$ introduced in Eq. 10 in order to derive the weighted mean of the explanatory variables on the
392 cluster level. Neglecting the support adjustment would deteriorate the coherence between explanatory
393 variables computed at the forest boundary and the corresponding local density that already includes
394 a potential boundary adjustment, thus introducing unnecessary noise to the model. The boundary
395 adjustment to the support also makes the sampling frame more consistent for the different data sources
396 (Section 5.3).

397 The ALS acquisition year (*ALSpyear*) was added as a categorical variable in order to account for the
398 time lag with the terrestrial survey as well as to help explain the heterogeneity in the data introduced
399 by the varying ALS quality. In 2008, a sensor error produced particularly poor ALS quality so the year
400 was divided accordingly into two factor levels, denoted 2008_1 and 2008. Furthermore, in order to
401 increase the number of observations per factor level the years 2006 and 2007 were pooled together and

⁴⁰² the same was done for 2012 and 2013. The result was nine factor levels denoted as 2002, 2003, 2007,
⁴⁰³ 2008_1, 2008, 2009, 2010, 2011 and 2012.

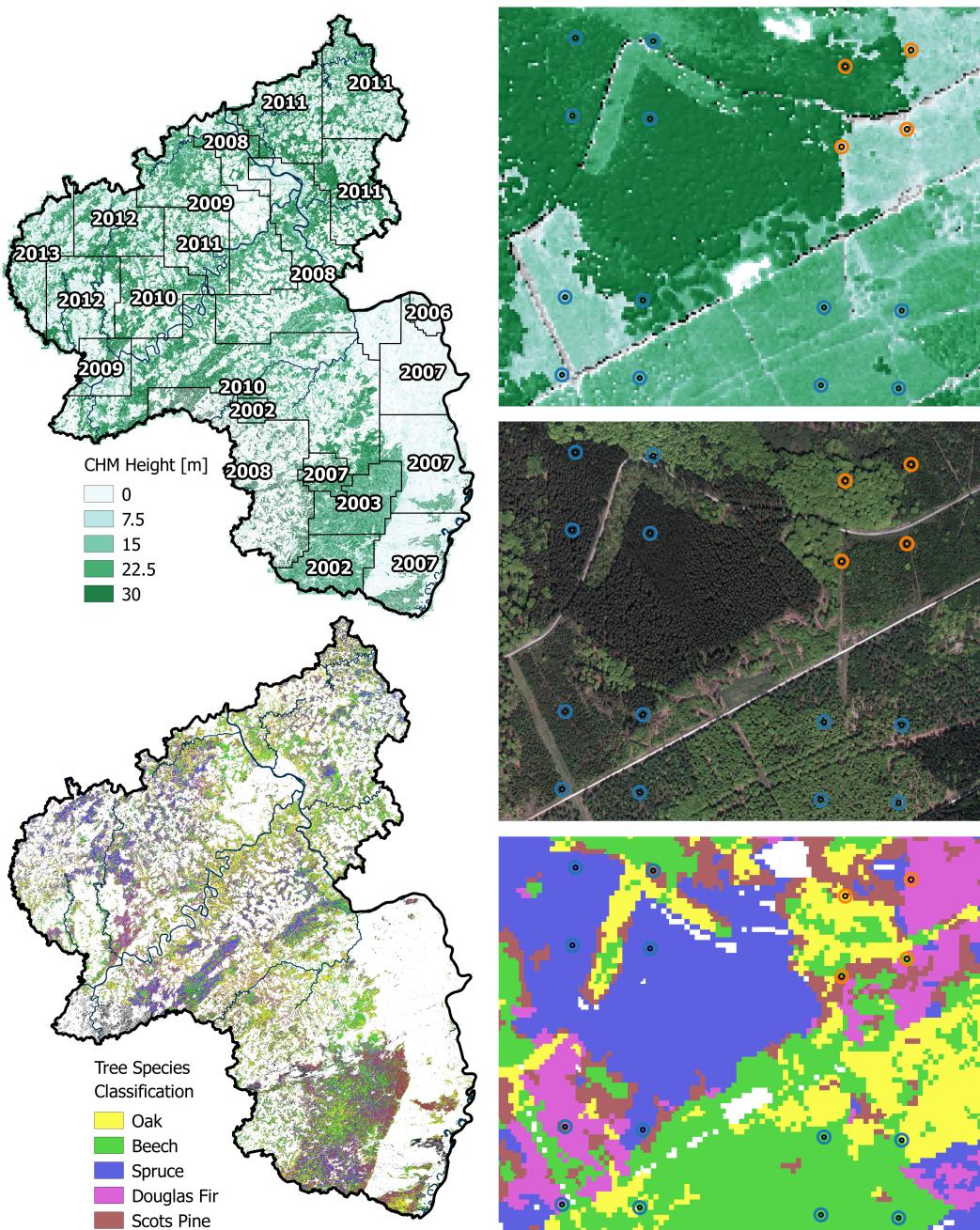


Figure 2. Left: CHM (top) and tree species classification map (bottom) available on the federal state level. Right: Magnified illustration of the supports used to derive the explanatory variables from the auxiliary data. From top to bottom: CHM, aerial image, tree species classification.

⁴⁰⁴ 5.5.2. Tree species classification map

⁴⁰⁵ The tree species map was used to predict the main tree species at each sample plot which served
⁴⁰⁶ as an additional categorical variable *treespecies* in the regression model. In the first step, one of the five
⁴⁰⁷ tree species was assigned to a sample location if 100% of the raster values within the edge-corrected
⁴⁰⁸ support were classified as that species. Otherwise, the sample location was assigned the value 'mixed'.
⁴⁰⁹ 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. The specific setting for the support size and the percentage threshold was found to be optimal in order to yield the best possible regression model precision when incorporating the *treespecies* variable as an additional predictor. 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 [45] 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 [46] of the *treespecies* variable after calibration. More details on the processing of the explanatory variables and identification of optimal parameter settings for their calculation are described in Hill *et al.* [41].

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	22.3	47.0	883	419
Douglas Fir	24.8	48.7	230	117
Oak	11.1	48.5	289	66
Spruce	53.2	61.1	651	566
Scots Pine	22.9	46.1	179	89
Mixed	84.5	64.5	3152	4127
Overall accuracy: 62.0 %			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.* [41] if more details are desired.

The model with highest adjusted R^2 and lowest RMSE was achieved using *meanheight*, *meanheight*², *stddev*, *ALyear* and *treespecies* as main effects, and including interaction terms between *meanheight* and *ALyear*, *stddev* and *ALyear*, *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 m³/ha and 33.6%). Using the ALS acquisition year as a categorical variable 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, this also led to a highly unbalanced data set when introducing the *treespecies* variable as an additional categorical predictor. For this reason, individual species modeling within each *ALyear* 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^2 s 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. For this reason, a joint evaluation of the estimation results is performed in 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:ALSpyear + meanheight:stddev + stddev:ALSpyear	full model	0.58 (0.48)	90.11 (139.22)	29.76 (45.98)
meanheight + stddev + meanheight ² + ALSyear + meanheight:ALSpyear + meanheight:stddev + stddev:ALSpyear	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 (*ALSpyear*). *Area_{ALSpyear}*: Area covered by ALS acquisition given in km². *n*: sample size of validation data. () give the respective values on the plot level.

<i>ALSpyear</i>	<i>Area_{ALSpyear}</i>	R^2	RMSE	RMSE%	<i>n</i>
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 treatment of outliers or leverage points, it can be advisable to remove such observations from the training data set that is used to determine the regression coefficients (Section 4.2) in order to minimize the residual variance for the entire terrestrial sample. However, it should be noted that such removal of observations does restrict the calculation of design-unbiased estimation to the PSMALL estimator, because the residuals still have to be derived for the full terrestrial sample in

order to ensure unbiased estimation. This would no longer be satisfied when using the EXTPSYNTH estimator, where the residual correction term is included in the regression coefficient (Section 4.2.3). For our particular case, we conducted an analysis of influential observations [47, 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 provide more justification for the inclusion of the *ALSyear* variable to mitigate the implied 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. Also the PSYNTH estimator could not be applied for 2 FR-units since $n_{1,G} < 2$. For better comparison, the descriptive summaries of point estimates and estimation errors for each estimator presented in Table 6 are always based on the same population of small area units, i.e. the 321 districts for which unbiased two-phase estimations were possible. All estimators could however be applied to all 45 FA units due to substantially larger sample sizes. 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 for the volume of growing stock per area unit [m^3/ha] 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 ($N_u=321$)	302.77	99.89	552.87	16.94	2.76	55.51
	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 ($N_u=321$)	305.56	197.47	444.29	4.13	2.56	18.07

6.2. Estimation errors

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 16.9%, 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 EXTPSYNTH estimator. 95% of the 321 FR units where PSMALL and EXTPSYNTH could be applied exhibited errors less than 20%. In comparison, the SRS estimates resulted in errors less than 32.3% for 95% of the 321 evaluated FR units.

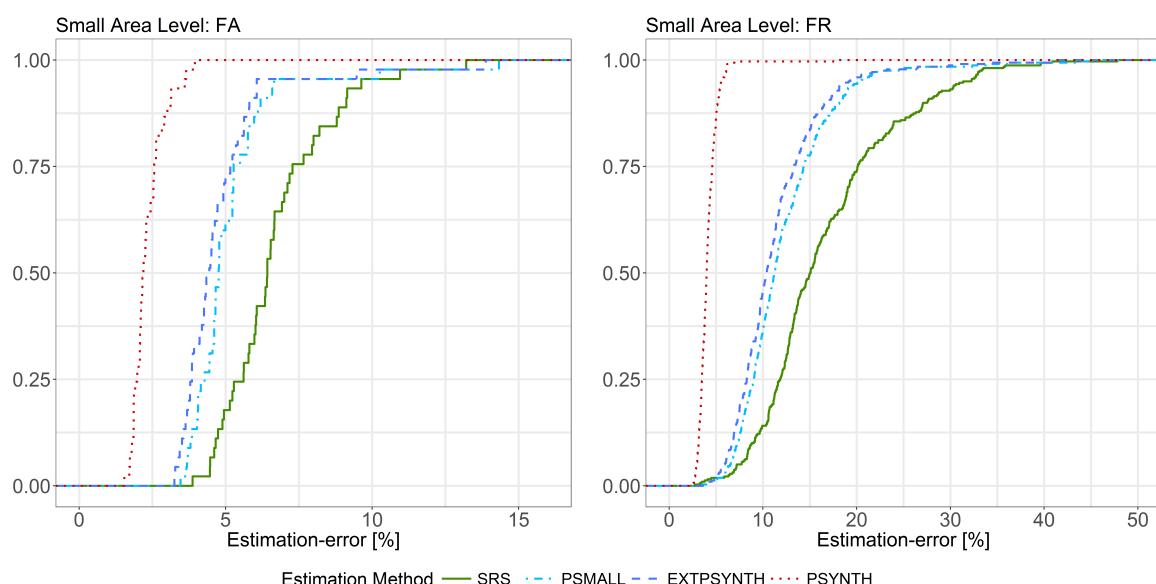


Figure 3. Cumulative distribution of estimation errors under SRS, PSMALL, EXTPSYNTH and the PSYNTH estimator. *Left:* Results for the 45 FA units. *Right:* Results for the 321 FR units.

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.

512 6.3. Comparison of PSMALL and EXTPSYNTH

513 Figure 3 reveals that the error distribution of PSMALL and EXTPSYNTH are very similar, with
 514 PSMALL showing marginally higher estimation errors. In order to investigate the differences between
 515 PSMALL and EXTPSYNTH, we compared the g-weight variances of both estimators for all 321 FR
 516 units (Fig. 4, left). As obvious, PSMALL yielded slightly larger variances for the vast majority of
 517 the estimates. As addressed in Section 4.2.3, one possible explanation for differences was the effect
 518 of one or more clusters not entirely being included in a small area unit, as this would constitute an
 519 assumption violation of the EXTPSYNTH estimator. This violation was actually observed in 155 of
 520 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 [48] 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 EXTPSYNTNTH 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 [35, 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 EXTPSYNTNTH 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 EXTPSYNTNTH 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 EXTPSYNTNTH 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 EXTPSYNTNTH variances of those small areas with violations and the EXTPSYNTNTH 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 EXTPSYNTNTH revealed that the variance differences did not lead to statistically significant point estimates.

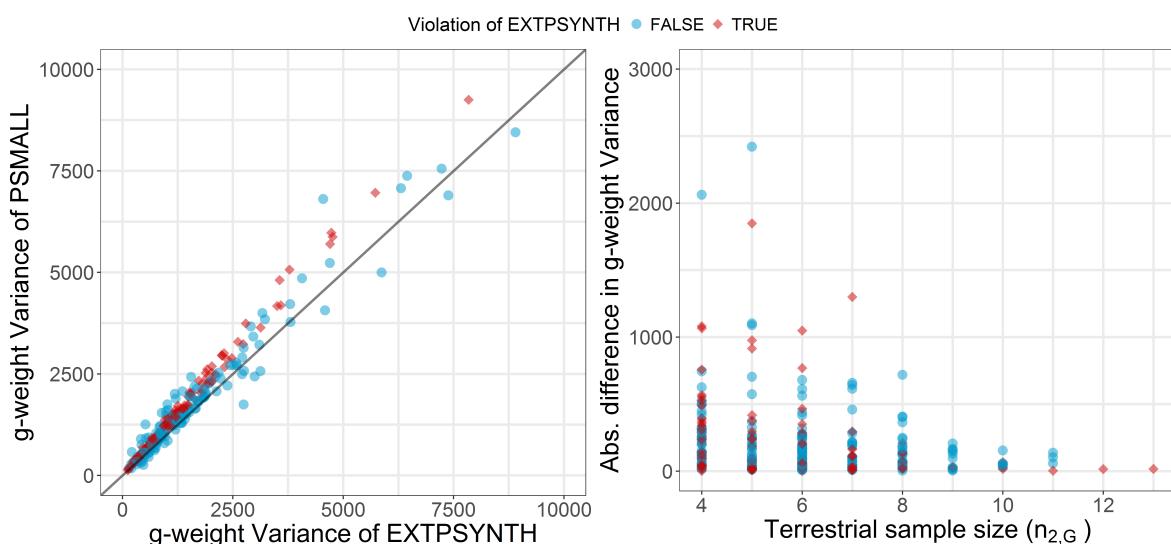


Figure 4. *Left:* Comparison of the g-weight variance between the PSMALL and the EXTPSYNTNTH estimator for the 321 FR units. *Right:* Difference in g-weight variance between the PSMALL and the EXTPSYNTNTH estimator in dependence of the terrestrial data ($n_{2,G}$) in the FR unit.

6.4. Variance reduction compared to SRS

The variance reduction relative to SRS for PSMALL and EXTPSYNTNTH 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, EXTPSYNTNTH) led to a variance reduction compared to SRS in all FA units. In 75% of the FA units, the EXTPSYNTNTH 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 EXTPSYNTNTH estimator (Section 6.3).

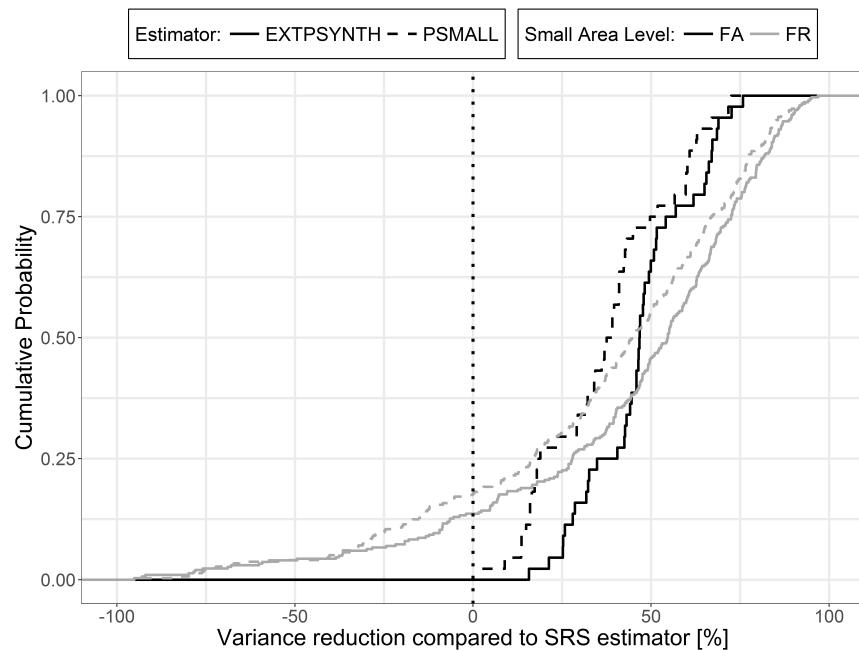


Figure 5. Cumulative distribution of variance reduction by the PSMALL and EXTPSYNTH compared to the SRS estimator for the 45 FA and 321 FR units.

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

550 Cases also occurred on the FR level where one or both two-phase estimators produced larger
 551 variance values than under the SRS estimator. This happened in 19% of the FR units under the
 552 EXTPSYNTH, and in 24% of the FR units under the PSMALL estimator. One possible reason for this
 553 was supposed to be a large residual variance due to a poor performance of the regression model
 554 within the small area unit. In order to investigate this hypothesis, we analyzed the three variance
 555 terms of the PSMALL estimator (Eq. 14b), i.e. the variance introduced by the uncertainty of the
 556 regression coefficients (term 1), the variance caused by estimating the auxiliary means (term 2), and
 557 the variance of the model residuals (term 3). In general, the residual term is expected to make the
 558 largest contribution to the overall variance since it's sample size is based on $n_{2,G}$ whereas the auxiliary
 559 term and the coefficient term are based on larger sample sizes, i.e. $n_{1,G}$ and n_2 respectively. Figure 6
 560 illustrates the share of the overall variance by the residual term of the PSMALL estimator scaled by
 561 the overall percentage reduction or increase of the variance compared to SRS for various small area
 562 sample sizes $n_{2,G}$. Not surprisingly, the residual term generally constitutes the dominating part of the
 563 PSMALL variance (around 84% on average). It has to be noted that such high residual term dominance
 564 does not necessarily indicate that the PSMALL variance will be disproportionately large (Figure 6, right).
 565 However, the vast majority of cases where the PSMALL variance was considerably larger than the
 566 SRS variance occurred where the residual term contributed over 75% to the overall PSMALL variance
 567 (Figure 6, left). Among those cases, the most pronounced were observed under small sample sizes

⁵⁶⁸ $n_{2,G} < 5$. Here, the average increase in variance compared to SRS of those FR units with $n_{2,G} = 4$ was
⁵⁶⁹ 272%, compared to 62% for FR units with $n_{2,G} > 4$. In contrast, the decreases in variance compared
⁵⁷⁰ to SRS (Figure 6, right) were much more homogeneous in magnitude and also independent of the
⁵⁷¹ terrestrial sample size. Since $n_{2,G}$ is the same for PSMALL and SRS, these observations imply that in
⁵⁷² the problematic small areas, the sum of square residuals for the regression model are likely larger than
⁵⁷³ the sum of square local densities for the clusters in $s_{2,G}$. This indicates the presence of outliers with
⁵⁷⁴ large residuals, which likely arise when there was forest loss after the ALS scanning but before the
⁵⁷⁵ terrestrial survey year.

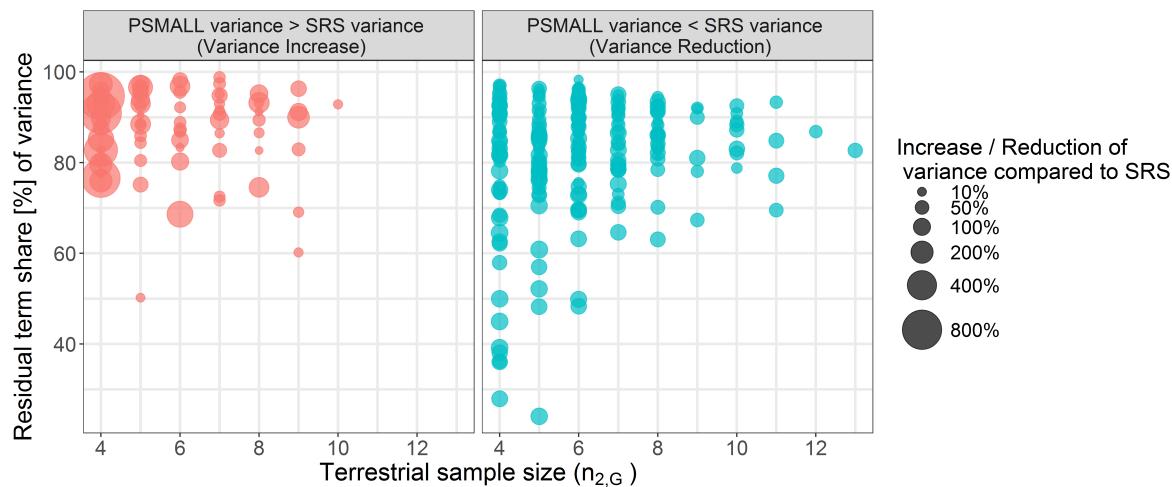


Figure 6. Share of the overall variance by the residual term of the PSMALL estimator for various small area sample sizes. Points are scaled by the overall percentage reduction/increase of the variance compared to SRS.

⁵⁷⁶ 7. Discussion

⁵⁷⁷ 7.1. Performance of estimators

⁵⁷⁸ With the objective of extending the use of the German NFI data to additional estimation on
⁵⁷⁹ small-scale management levels, we evaluated the performance of design-based small area regression
⁵⁸⁰ estimators with respect to their suitability for future operational large scale application. For this reason,
⁵⁸¹ we conducted a case study in the German federal state of Rhineland-Palatinate where we applied the
⁵⁸² SRS, the PSMALL and the EXTPSYNTH estimators to produce estimates of the mean timber volume on
⁵⁸³ two forest management levels over the entire federal state area, comprising 45 and 405 small area units
⁵⁸⁴ respectively. In order to assess and compare the performance of the estimators, it was of particular
⁵⁸⁵ interest to gather information about the magnitudes of estimation precision they can provide.

⁵⁸⁶ Our study showed that on both small area levels, the PSMALL and the EXTPSYNTH estimators
⁵⁸⁷ generally led to a substantial reduction in estimation error compared to the standard one-phase
⁵⁸⁸ SRS estimator. On the upper management level (FA districts), PSMALL and EXTPSYNTH produced
⁵⁸⁹ estimation errors smaller than 5% for 73% of the small areas compared to only 17% under the one-phase
⁵⁹⁰ SRS estimator. It could be argued that the majority of the FA units comprised a sample size (46 clusters
⁵⁹¹ on average) that would also have allowed to build an individual prediction model for each FA unit and
⁵⁹² apply direct estimation, which can be more efficient. The main reason why this was not considered in
⁵⁹³ the present study was the large number of parameters that would have to be fitted in the individual
⁵⁹⁴ prediction models resulting from the pronounced heterogeneity in the auxiliary data (multiple tree
⁵⁹⁵ species and ALS acquisition years in each FA unit). In this case, the strategy to 'borrow strength' from
⁵⁹⁶ the entire inventory domain was preferred to direct estimation in order to avoid overfitting and the
⁵⁹⁷ implied risk of unstable global estimation. Even in case the above mentioned problem of overfitting is

598 not raised, individual model building for a large number of small area units can be time-consuming
599 and small area estimation can be a cost-saving alternative - however, possibly the cost of more efficient
600 global estimation.

601 The level of precision of the FA-level could not be achieved on the lower management level
602 (FR districts) primarily due to substantially smaller terrestrial sample sizes. However, in 95% of the
603 FR units, the estimation errors could be limited to 20% compared to 40% under SRS. A pairwise
604 comparison of the confidence intervals revealed that the estimators did not produce significantly
605 different point estimates. The much smaller estimation errors of the PSYNTH estimator reflected the
606 fact that it does not try to correct for potential bias in the point estimate which can lead to overly
607 optimistic estimation errors and confidence intervals. One should thus prefer the unbiased estimates
608 of PSMALL or EXTPSYNTH whenever their calculation is possible.

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

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

638 A commonly raised critic on the proposed design-unbiased estimators are their asymptotic
639 properties, i.e. the validity of the confidence coverage rates is only ensured if the sample size in a small
640 area is sufficiently large. Giving a generally valid sample size for the asymptotic validity range is,
641 unfortunately, infeasible due to the dependency on the heterogeneity of the underlying population
642 which is per se unknown. However, simulation studies for simple random sampling presented in
643 Mandallaz *et al.* [25] suggested that a minimum sample size of 6 within a small area is sufficient
644 to ensure the nominal coverage rates of the confidence intervals for PSMALL and EXTPSYNTH.
645 Re-evaluating the same simulation example recently confirmed the same results for cluster sampling.

646 7.2. Auxiliary data

647 The auxiliary data used in our study were derived from two remote sensing sources, i.e. an ALS
648 canopy height model and a tree species classification map. Likewise in many similar studies, the ALS
649 mean canopy height proved to be the explanatory variable with highest predictive power. However,
650 the large time-gaps of up to 10 years between the ALS acquisition and the terrestrial survey date caused
651 the substantial introduction of artificial noise in the data. Whereas a post-stratification to the ALS
652 acquisition years was an effective means to counteract the implied residual inflation, several leverage
653 points were unambiguously caused by the temporal asynchronicity. Undetectable forest loss during
654 the gap between the ALS acquisition and the NFI was also likely a cause for high residual variance
655 in some small area units compared to the terrestrial data variance, which subsequently led to higher
656 variances than the SRS estimator. As opposed to the ALS data, the availability of a country-wide tree
657 species classification map has yet been unique among all German federal states. Whereas the study of
658 Hill *et al.* [41] already showed that the tree species information was able to improve the model fit, it has
659 yet not been used to its full potential. One reason for this was the impossibility of modeling individual
660 tree species within each ALS acquisition year, which would add further explanatory power. Another
661 reason was the lack of available satellite data for classification in some parts of the country, which
662 led to missing values in the inventory data and restricted 19 FR units to a simpler regression model.
663 Promising steps with respect to more up-to-date canopy height information have already been made, as
664 the topographic survey institution of RLP will from this year on provide a country-wide canopy height
665 model derived from aerial imagery acquisitions. These campaigns will in the future be conducted in a
666 two-year period and allow to derive canopy height information matching the dates of terrestrial forest
667 inventories. A study of Kirchhoefer *et al.* [49] recently indicated that similar model performance for
668 German NFI data can be achieved using such imagery-based canopy height models. Additionally, the
669 improved coverage and repetition rate of the Sentinel-2 satellite [50] will allow to produce annually
670 updated tree species classification maps. We consider these alternative auxiliary data sources to also
671 solve the problem of missing explanatory variables at inventory plots. One could also make use of
672 the exhaustive information within the two-phase estimators by using the true auxiliary means [25,30],
673 which could further decrease estimation errors. Previous studies of Mandallaz *et al.* [25] however
674 showed that given a reasonable large sample size of the first phase, the differences in the estimation
675 error are usually small. With respect to the substantial improvements in the temporal synchronicity
676 between auxiliary and terrestrial inventory data, we consider the demonstrated double-sampling
677 approach also to be very efficient for the estimation of change [51].

678 8. Conclusion

679 The study led to two major conclusions: (1) the EXTPSYNTH and PSMALL estimator generally
680 achieved substantially smaller estimation errors on the two investigated forest district levels compared
681 to the SRS estimator. Thus, the demonstrated small area estimation procedure constitutes a major
682 contribution to an additional use of the German NFI data for estimation below the federal state
683 level. Further close cooperation with the forest authorities is crucial to evaluate whether the achieved
684 error levels are already sufficient enough in order to support forest planning decisions. A first
685 study will concentrate on testing the EXTPSYNTH and PSMALL confidence intervals as a validation
686 source for the stand-wise inventories. (2) Despite the quality restrictions, the ALS data and the tree
687 species map were found to be well suited to model the mean timber volume on the plot and cluster
688 level. With the prospect of more frequently updated aerial canopy height models and tree species
689 maps, the two data sources will become even more attractive to be used as an integral part of future
690 operational applications. The improving availability of remote sensing data will also allow to extent
691 the demonstrated estimation procedure to the estimation of change. We consider this to be one of the
692 next milestones towards a future operational use of the demonstrated small area estimation procedure.

Acknowledgments: We want to express our gratitude to Prof. H. Heinemann (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 Institute 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.

$$R^2 = \frac{\sum_{x \in s_2} \left(\frac{M(x)}{M_2} \right)^2 \left(\hat{Y}_c(x) - \hat{Y}_c \right)^2}{\sum_{x \in s_2} \left(\frac{M(x)}{M_2} \right)^2 \left(Y_c(x) - \hat{Y}_c \right)^2}$$

$Y_c(x)$ and $\hat{Y}_c(x)$ are the predicted and observed local densities on the cluster level calculated according to Equations 2 and 12. \hat{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.

$$RMSE = \sqrt{\frac{1}{n_2} \sum_{x \in s_2} \left(\frac{M(x)}{M_2} \right)^2 \left(\hat{Y}_c(x) - Y_c(x) \right)^2}$$

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

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

Note that the weights $\frac{M(x)}{M_2} \equiv 1$ if the number of plots per cluster is constant.

References

1. Polley, H.; Schmitz, F.; Hennig, P.; Kroher, 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 study of European forests. *Annals of Forest Science* **2016**, *73*, 857–869.
4. Böckmann, T.; Saborowski, J.; Dahm, S.; Nagel, J.; Spellmann, H. Neukonzeption und Weiterentwicklung der Forsteinrichtung in Niedersachsen. *Forst und Holz (Germany)* **1998**.
5. 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.

- 728 6. von Lüpke, N. Approaches for the Optimisation of Double Sampling for Stratification in Repeated Forest
729 Inventories. PhD thesis, University of Göttingen, 2013.
- 730 7. De Vries, P.G. *Sampling theory for forest inventory: a teach-yourself course*; Springer, 1986.
- 731 8. Cochran, W.G. *Sampling techniques*, wiley series in probability and mathematical statistics: applied
732 probability and statistics ed.; Wiley, 1977.
- 733 9. Särndal, C.E.; Swensson, B.; Wretman, J. *Model assisted survey sampling*; Springer Science & Business Media,
734 2003.
- 735 10. Gregoire, T.G.; Valentine, H.T. *Sampling strategies for natural resources and the environment*; CRC Press, 2007.
- 736 11. Köhl, M.; Magnussen, S.S.; Marchetti, M. *Sampling methods, remote sensing and GIS multiresource forest
737 inventory*; Springer Science & Business Media, 2006.
- 738 12. Mandallaz, D. *Sampling techniques for forest inventories*; CRC Press, 2008.
- 739 13. Gillis, M.; Boudewyn, P.; Power, K.; Russo, G. Canada. In *National Forest Inventories - Pathways for Common
740 Reporting*; Springer, 2010; chapter 4, pp. 97–111.
- 741 14. Chojnacky, D.C. Double sampling for stratification: a forest inventory application in the Interior West.
742 Technical report, US Department of Agriculture, Forest Service, Rocky Mountain Research Station Ogden,
743 UT, 1998.
- 744 15. Lanz, A.; Brändli, U.B.; Brassel, P.; Ginzler, C.; Kaufmann, E.; Thürig, E. Switzerland. In *National Forest
745 Inventories - Pathways for Common Reporting*; Springer, 2010; chapter 36, p. 555–565.
- 746 16. Gasparini, P.; Tosi, V.; DiCosmo, L. Italy. In *National Forest Inventories - Pathways for Common Reporting*;
747 Springer, 2010; chapter 19, p. 311–331.
- 748 17. Saborowski, J.; Marx, A.; Nagel, J.; Böckmann, T. Double sampling for stratification in periodic
749 inventories—Infinite population approach. *Forest ecology and management* **2010**, *260*, 1886–1895.
- 750 18. Grafström, A.; Schnell, S.; Saarela, S.; Hubbell, S.; Condit, R. The continuous population approach to forest
751 inventories and use of information in the design. *Environmetrics* **2017**.
- 752 19. Massey, A.; Mandallaz, D.; Lanz, A. Integrating remote sensing and past inventory data under the new
753 annual design of the Swiss National Forest Inventory using three-phase design-based regression estimation.
754 *Canadian Journal of Forest Research* **2014**, *44*, 1177–1186.
- 755 20. Mandallaz, D. A three-phase sampling extension of the generalized regression estimator with partially
756 exhaustive information. *Canadian Journal of Forest Research* **2013**, *44*, 383–388.
- 757 21. Rao, J.N. *Small-Area Estimation*; Wiley Online Library, 2015.
- 758 22. Breidenbach, J.; Astrup, R. Small area estimation of forest attributes in the Norwegian National Forest
759 Inventory. *European Journal of Forest Research* **2012**, *131*, 1255–1267.
- 760 23. Goerndt, M.E.; Monleon, V.J.; Temesgen, H. A comparison of small-area estimation techniques to estimate
761 selected stand attributes using LiDAR-derived auxiliary variables. *Canadian journal of forest research* **2011**,
762 *41*, 1189–1201.
- 763 24. Steinmann, K.; Mandallaz, D.; Ginzler, C.; Lanz, A. Small area estimations of proportion of forest and
764 timber volume combining Lidar data and stereo aerial images with terrestrial data. *Scandinavian journal of
765 forest research* **2013**, *28*, 373–385.
- 766 25. Mandallaz, D.; Breschan, J.; Hill, A. New regression estimators in forest inventories with two-phase
767 sampling and partially exhaustive information: a design-based monte carlo approach with applications to
768 small-area estimation. *Canadian Journal of Forest Research* **2013**, *43*, 1023–1031.
- 769 26. Koch, B. Status and future of laser scanning, synthetic aperture radar and hyperspectral remote sensing
770 data for forest biomass assessment. *ISPRS Journal of Photogrammetry and Remote Sensing* **2010**, *65*, 581–590.
- 771 27. Naesset, E. Area-Based Inventory in Norway – From Innovation to an Operational Reality. In *Forest
772 Applications of Airborne Laser Scanning - Concepts and Case Studies*; Springer, 2014; chapter 11, pp. 216–240.
- 773 28. Magnussen, S.; Mauro, F.; Breidenbach, J.; Lanz, A.; Kändler, G. Area-level analysis of forest inventory
774 variables. *European Journal of Forest Research* **2017**, pp. 1–17.
- 775 29. Magnussen, S.; Mandallaz, D.; Breidenbach, J.; Lanz, A.; Ginzler, C. National forest inventories in the
776 service of small area estimation of stem volume. *Canadian Journal of Forest Research* **2014**, *44*, 1079–1090.
- 777 30. Mandallaz, D. Design-based properties of some small-area estimators in forest inventory with two-phase
778 sampling. *Canadian Journal of Forest Research* **2013**, *43*, 441–449.
- 779 31. Bundesministerium für Ernährung, L.u.V. Aufnahmeanweisung für die dritte Bundeswaldinventur BWI3
780 (2011 - 2012), 2011.

- 781 32. Bitterlich, W. *The relascope idea. Relative measurements in forestry.*; Commonwealth Agricultural Bureaux,
782 1984.
- 783 33. Kublin, E.; Breidenbach, J.; Kändler, G. A flexible stem taper and volume prediction method based on
784 mixed-effects B-spline regression. *European journal of forest research* **2013**, *132*, 983–997.
- 785 34. Schmitz, F.; Polley, H.; Hennig, P.; Dunger, K.; Schwitzgebel, F. Die zweite Bundeswaldinventur -
786 BWI2: Inventur- und Auswertmethoden, Bundesministerium für Ernährung, Land- Wirtschaft und
787 Verbraucherschutz (Hrsg), 2008.
- 788 35. Mandallaz, D.; Hill, A.; Massey, A. Design-based properties of some small-area estimators in forest
789 inventory with two-phase sampling - revised version. Technical report, Department of Environmental
790 Systems Science, ETH Zurich, 2016.
- 791 36. Hill, A.; Massey, A. *forestinventory: Design-Based Global and Small-Area Estimations for Multiphase Forest
792 Inventories. R package version 0.3.1*, 2017.
- 793 37. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical
794 Computing, Vienna, Austria, 2018.
- 795 38. Gauer, J.; Aldinger, E. Waldökologische Naturräume Deutschlands-Wuchsgebiete. *Mitteilungen des Vereins
796 für Forstliche Standortskunde und Forstpflanzenzüchtung* **2005**, *43*, 281–288.
- 797 39. LWaldG. *Landeswaldgesetz Rheinland-Pfalz (Forest Act Rhineland-Palatinate)*, 2000. Rhineland-Palatinate,
798 Germany.
- 799 40. Lamprecht, S.; Hill, A.; Stoffels, J.; Udelhoven, T. A Machine Learning Method for Co-Registration and
800 Individual Tree Matching of Forest Inventory and Airborne Laser Scanning Data. *Remote Sensing* **2017**, *9*.
- 801 41. Hill, A.; Buddenbaum, H.; Mandallaz, D. Combining canopy height and tree species map information for
802 large-scale timber volume estimations under strong heterogeneity of auxiliary data and variable sample
803 plot sizes. *European Journal of Forest Research* **2018**.
- 804 42. Stoffels, J.; Hill, J.; Sachtleber, T.; Mader, S.; Buddenbaum, H.; Stern, O.; Langshausen, J.; Dietz, J.;
805 Ontrup, G. Satellite-Based Derivation of High-Resolution Forest Information Layers for Operational Forest
806 Management. *Forests* **2015**, *6*, 1982–2013.
- 807 43. Stoffels, J.; Mader, S.; Hill, J.; Werner, W.; Ontrup, G. Satellite-based stand-wise forest cover type mapping
808 using a spatially adaptive classification approach. *European journal of forest research* **2012**, *131*, 1071–1089.
- 809 44. White, J.C.; Coops, N.C.; Wulder, M.A.; Vastaranta, M.; Hilker, T.; Tompalski, P. Remote sensing
810 technologies for enhancing forest inventories: A review. *Canadian Journal of Remote Sensing* **2016**,
811 *42*, 619–641.
- 812 45. Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5–32.
- 813 46. Congalton, R.G.; Green, K. *Assessing the accuracy of remotely sensed data: principles and practices*; CRC press,
814 2008.
- 815 47. Fahrmeir, L.; Kneib, T.; Lang, S.; Marx, B. *Regression: models, methods and applications*; Springer Science &
816 Business Media, 2013.
- 817 48. Wilcoxon, F.; Katti, S.; Wilcox, R.A. Critical values and probability levels for the Wilcoxon rank sum test
818 and the Wilcoxon signed rank test. *Selected tables in mathematical statistics* **1970**, *1*, 171–259.
- 819 49. Kirchhoefer, M.; Schumacher, J.; Adler, P.; Kändler, G. Considerations towards a Novel Approach for
820 Integrating Angle-Count Sampling Data in Remote Sensing Based Forest Inventories. *Forests* **2017**, *8*, 239.
- 821 50. ESA. Sentinel-2 earth observation mission, 2017. Accessed: 2017-03-29.
- 822 51. Massey, A.; Mandallaz, D. Design-based regression estimation of net change for forest inventories. *Canadian
823 Journal of Forest Research* **2015**, *45*, 1775–1784, [<https://doi.org/10.1139/cjfr-2015-0266>].