

Article Draft

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

Large scale application to the federal state of Rhineland-Palatinate

Andreas Hill^{*†}, Daniel Mandallaz^{*}, Joachim Langshausen^{**}

^{*}Department of Environmental Systems Science, ETH Zurich

^{**}Forest Service Rhineland-Palatinate

Corresponding author[†]: andreas.hill@usys.ethz.ch
Co-authors: daniel.mandallaz@env.ethz.ch, joachim.langshausen@wald-rlp.de

Publication date: February 8, 2018

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 on the plot level. 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, double sampling for regression within strata, cluster sampling, LiDAR canopy height model, tree species classification

Contents

1	Introduction	1
2	Terrestrial sampling design of the German NFI	3
3	Double sampling in the infinite population approach	3
4	Estimators	4
4.1	Design-based SRS estimator for cluster sampling	4
4.2	Design-based small area regression estimators for cluster sampling	6
4.2.1	Pseudo Small Area Estimator (PSMALL)	7
4.2.2	Pseudo Synthetic Estimator (PSYNTH)	8
4.2.3	Extended Pseudo Synthetic Estimator (EXTPSYNTH)	8
4.3	Measures of estimation accuracy	9
5	Case study	10
5.1	Study area and small area units	10
5.2	Terrestrial sample	10
5.3	Extension to double sampling design	11
5.4	Auxiliary data	12
5.4.1	LiDAR canopy height model	12
5.4.2	Tree species map	13
5.5	Calculation of the explanatory variables	15
5.5.1	ALS canopy height model	15
5.5.2	Tree species map	15
5.6	Regression Model	16
6	Results	20
6.1	General estimation results	20
6.2	Estimation errors	20
6.3	Comparison of PSMALL and EXTPSYNTH	22
6.4	Reduction of SRS variance by PSMALL and EXTPSYNTH	23

List of Figures

1	<i>Left:</i> Study Area with delineated FA forest management units. <i>Right:</i> Example for each of the three management units (from top to bottom): FA, FR and forest stand unit with overlayed the extended double sampling cluster design. <i>Green:</i> Forest stand polygon layer defining the forest area of this study.	11
2	Left: CHM (top) and tree species classification map (bottom) available on the federal state level were used as auxiliary information. Right: Magnified illustration of the supports used to derive the explanatory variables from the auxiliary data.	14
3	Cumulative distribution of estimation errors under the simple random sampling (SRS), the pseudo small (PSMALL), the extended pseudo synthetic (EXTPSYNTH) and the pseudo synthetic (PSYNTH) estimator. <i>Left:</i> Results for the 45 FA units. <i>Right:</i> Results for the 388 (SRS), 321 (PSMALL / EXTPSYNTH) and 403 (PSYNTH) FR units.	21
4	<i>Left:</i> Comparison of the g-weight variance between the PSMALL and the EXTPSYNTH estimator for the 321 FR units. <i>Right:</i> Difference in g-weight variance between the PSMALL and the EXTPSYNTH estimator in dependence of the terrestrial data ($n2G$) in the FR unit.	22
5	Cumulative distribution of variance reduction by the PSMALL and EXTPSYNTH compared to the SRS estimator for the 45 FA and 321 FR units.	23
6	25
7	25

List of Tables

1	Descriptive statistics of the forest observed on NFI sample plots located within communal and state forest area (n=5791).	12
2	Sample size for each phase in entire study area. $n_{\{1,2\},plots}$: number of plots. $n_{\{1,2\}}$: number of cluster	13
3	Classification accuracies of the <i>treespecies</i> variable before and after calibration. n_{ref} : number of terrestrial reference plots. n_{class} : number of classified plots.	16
4	Accuracy metrics for the two OLS regression models. Interaction terms are indicated by ':'. () give the respective values on the cluster level.	17
5	R^2 , RMSE and RMSE% of the full regression model within ALS acquisition year strata (<i>ALSpyear</i>). <i>Area_{ALSpyear}</i> : Area covered by ALS acquisition given in km ² . n : number of validation data. () give the respective values on the cluster level.	19
6	Descriptive summary of point estimates and estimation errors on the two forest district levels. N_u : number of evaluated small area units.	20
7	Descriptive summary of SRS variance reduction and relative efficiencies on the two forest district levels. N_u : number of evaluated small area units.	24

1 Introduction

The German National Forest Inventory (NFI) provides reliable evidence-based and accurate information of the current state and the development of Germanys forest over time. The NFI thereby has to satisfy various information needs and amongst others reports 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 United Nations Framework Convention on Climate Change (UNFCCC) on the international level ([Polley et al, 2010](#)). At the current time, the inventory design of the German NFI solely rests upon a terrestrial cluster inventory that is carried out at sample locations systematically distributed over the entire forest state area of Germany. As this implies covering a large area of 114'191 ha ([Thünen-Institut, 2014](#)), the sample size has been chosen according to satisfy high estimation accuracies for forest attributes on the national and federal state level. This however leads to very low sampling intensities and consequently, 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 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 in operational forest planning on forest district management levels. In most German federal states, management strategies are thus still based on expert judgements from time-consuming standwise inventories (SFI), which are prone to systematic deviations [Kuliešis et al \(2016\)](#) 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 base their regional management strategies on quantitative and accurate information ([Böckmann et al, 1998](#)). 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 ([von Lüpke, 2013](#)). One method which has proven to be efficient is double-or two-phase sampling ([Särndal et al, 2003; Gregoire and Valentine, 2007; Köhl et al, 2006; Mandallaz, 2008](#)). Double sampling incorporates inexpensive auxiliary information and can be used to either increase estimation precision under given terrestrial sample size, or maintain estimation precision under reduced terrestrial sample size. A double sampling for stratification procedure has e.g. been used in the FDI of Lower Saxony ([Saborowski et al, 2010](#)), and [Grafström et al \(2017\)](#) lately illustrated how to use the auxiliary information to determine optimised balanced terrestrial sample designs. Recent studies have lately extended double-sampling to triple-sampling estimation methods using auxiliary information in two different sampling intensities. An example can be found in [von Lüpke et al \(2012\)](#) who illustrated an extension of the existing two-phase FDI 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. An other example is [Massey et al \(2014\)](#) who developed a triple-sampling extension based on the ideas of [Mandallaz \(2013b\)](#) for the Swiss NFI that can significantly reduce the increase in estimation uncertainty caused by the new annual inventory design.

1. INTRODUCTION

Two-phase and three-phase samplings techniques have also been used in the service of small area estimation (SAE). SAE techniques particular address the situation where the number of samples within a subunit, so-called 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 ([Köhl et al, 2006](#)) originally comes from official statistics. A commonly applied SAE method is thereby known as indirect estimation ([Rao, 2015](#)), where statistical models are used to convert auxiliary information into predictions of the target variable that is rarely or not available in the small area. The statistical models are thereby often developed by "borrowing strength" from data outside the small area. There are numerous applications of SAE in forestry ([Breidenbach and Astrup, 2012](#); [Goerndt et al, 2011](#); [Steinmann et al, 2013](#); [Mandallaz et al, 2013](#)), and most of the studies use unit-level models, i.e. the statistical models are fitted using data from inventory plots. Especially unit-level models for timber volume estimation under the use of various remote sensing data have been intensively investigated with respect to timber volume prediction ([Koch, 2010](#); [Naesset, 2014](#)). There are also few studies that have investigated area level-models, where the auxiliary information is only provided on the SA-level ([Magnussen et al, 2017](#)). Some studies have illustrated that even NFI data of low sampling densities can be used in small area estimation procedures to provide estimations of acceptable accuracy on much smaller management levels. One example is [Breidenbach and Astrup \(2012\)](#) who used data from the Norwegian NFI for SAE estimations of standing timber volume for 14 municipalities where the number of NFI samples within these areas were between 1 and 35. Estimation errors under the applied model-based and design-based SAE estimators turned to be markedly smaller than achieved under simple random sampling (SRS). Another example is [Magnussen et al \(2014\)](#) who recently used the Swiss NFI data for SAE estimation of timber volume within 108 Swiss forest districts with sample sizes between 9 and 206. Despite these promising results, a similar study in Germany using the German NFI data for SAE estimation has not yet been conducted.

The aim of this study was to investigate whether the German NFI data can provide acceptable estimation precision on two forest district levels when incorporated in 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 \(2013a\)](#); [Mandallaz et al \(2013\)](#). Auxiliary data were obtained from a countrywide airborne Laser scanning (ALS) canopy height model (CHM) and a tree species classification map and used for regression within tree species strata. The estimation accuracies were compared to those achieved under SRS sampling. The chosen double-sampling estimators were favoured for several reasons: **(i)** the design-based frame considerably relaxes requirements on the regression model which seemed appropriate facing severe quality restrictions in the ALS data; **(ii)** the estimators can consider *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 the design-dependency of the regression coefficients on the sample (*internal model ap-*

proach) and is also robust to heteroscedasticity of the model residuals.

The results from this study were considered to provide valuable information whether the suggested procedure a) might be a cost-saving alternative to a regional FDI and b) can be used as a reliable validation for SDI data. A secondary objective was to address the potential effects of auxiliary data quality on estimation precision, and to identify challenges when transferring the existing German NFI design into the suggested double sampling estimation procedure.

2 Terrestrial sampling design of the German NFI

The German National Forest Inventory (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 the years 2011 and 2012. While information was originally gathered at 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 maximal four sample locations (also referred to as *sample plots*) that are arranged in a square (so called *cluster*) with a side length of 150 metres (Picture). The number of plots per cluster can however vary between 1 and 4 depending on forest/non-forest decisions by the field crews on the individual plot level ([Bundesministerium für Ernährung, 2011](#)). In the field survey of the BWI3, sample trees for timber volume estimation are selected according to the angle count sampling technique ([Bitterlich, 1984](#)), 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 tree-individual inclusion circle ([Bundesministerium für Ernährung, 2011](#)). 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 calculate a volume estimation on the tree level. These volume estimations 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 ([Kublin et al, 2013](#)).

3 Double sampling in the infinite population approach

The estimators used in this study have been proposed by ([Mandallaz, 2013a; Mandallaz et al, 2013](#)) and build upon the so called infinite population approach (IPA) in order to bridge the inventory procedure and the derived information to the mathematics behind the estimators. Therefore, we shall first provide a short introduction into this general estimation frame. We start by assuming that the population P of trees $i \in 1, 2, \dots, N$ within a forest of interest F is exactly defined, and each tree i has a directly or indirectly observable response variable Y_i (e.g., its timber volume) that allows to specify the population mean Y (e.g., the average timber volume per unit area) over F . If a full census of all tree population individuals is not possible, Y has to

be estimated by the conduction of an inventory. The infinite population approach assumes that the spatial distribution of the local density $Y(x)$ (e.g., the timber volume per unit area) at each point or location x in the forest F is given by a fixed (i.e. non stochastic) piecewise constant function. The population mean Y is thus mathematically equivalent to the integral of the density function surface divided by the forest area $\lambda(F)$, i.e. $Y = \frac{1}{\lambda(F)} \sum_{i=1}^N Y_i = \frac{1}{\lambda(F)} \int_F Y(x) dx$, and thus the population mean Y corresponds to a spatial mean. Since the local density function is in practice always unknown, one estimates Y by collecting a sample s_2 of all local density values by the conduction of a terrestrial inventory at n_2 uniformly and independently distributed sample points over F . This procedure is often referred to as *one-phase sampling* (OPS). Opposed 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 $\mathbf{Z}(x) = (z(x)_1, z(x)_2, \dots, z(x)_p)^\top$ 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.

4 Estimators

4.1 Design-based SRS estimator for cluster sampling

The simple random sampling (SRS) estimator for cluster sampling constitutes the *status quo* that is currently applied under the existing one-phase sampling design of the German NFI in order to obtain a point and variance estimate 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 ([Schmitz et al, 2008](#)) in the form given in [Mandallaz \(2008\)](#); [Mandallaz et al \(2016\)](#).

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, \dots, x_M are created 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 of plots per cluster, $M(x)$, is a random variable due to the uniform distribution of x_l ($l = 1, \dots, M$) in the forest F and forest/non-forest decision for each sample location x_l :

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

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

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

The local density $Y(x_l)$ on individual sample plot level was calculated according to the description in [Mandallaz \(2008\)](#), which can be rewritten for angle-count sampling technique applied in the BWI3. The general form of $Y(x)$ in [Mandallaz \(2008\)](#) is given as the Horwitz-Thompson estimator

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

where Y_i is in our case the predicted 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:

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

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

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

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:

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

Using 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 ([Schmitz et al, 2008](#)):

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

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{Y}_c and its estimated variance $\hat{\mathbb{V}}(\hat{Y}_c)$ for any given spatial unit for which $n_2 \geq 2$ (n_2 denoting the number of sample units, i.e. clusters):

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

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

4.2 Design-based small area regression estimators for cluster sampling

All three considered small area estimators have in common that they use ordinary least square (OLS) regression models to produce the predictions of the local density $Y_c(x)$ directly on the cluster level c . We consider the *internal model approach*, where 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:

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

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

$\mathbf{Z}_c(x)$ is the 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 support-area within the forest F used to derive the explanatory variables from the raw auxiliary information.

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

The estimated design-based variance-covariance matrix $\hat{\Sigma}_{\hat{\boldsymbol{\beta}}_{s_2}}$ accounts for the fact that the regression model is internal by reflecting the dependency of the estimated regression coefficients on the realized sample s_2 . It is defined as

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

with $\hat{R}_c = Y_c(x) - \mathbf{Z}_c^\top(x) \hat{\boldsymbol{\beta}}_{c,s_2}$ 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 [Mandalaz \(2013a\)](#); [Mandalaz et al \(2016, 2013\)](#) if the reader requires additional details or proofs. The estimators have also been implemented in the R-package *forestinventory* ([Hill](#)

and [Massey, 2017](#)) 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 $\hat{\beta}_{c,s_2}$ and $\hat{\Sigma}_{\hat{\beta}_{c,s_2}}$ which are always based on the entire sample s_2 . We thus first define the following quantities on the small area level:

$$\hat{\mathbf{Z}}_{c,G} = \frac{\sum_{x \in s_{1,G}} M_G(x) \mathbf{Z}_{c,G}(x)}{\sum_{x \in s_{1,G}} M_G(x)} \quad \text{where } \mathbf{Z}_{c,G}(x) = \frac{\sum_{l=1}^L I_G(x_l) \mathbf{Z}(x_l)}{M_G(x)} \quad (12a)$$

$$Y_{c,G}(x) = \frac{\sum_{l=1}^L I_G(x_l) \mathbf{Y}(x_l)}{M_G(x)} \quad \text{and } \hat{Y}_{c,G}(x) = \hat{\mathbf{Z}}_{c,G}^\top \hat{\beta}_{c,s_2} \quad (12b)$$

$$\bar{\hat{R}}_{2,G} = \frac{\sum_{x \in s_{2,G}} M_G(x) \hat{R}_{c,G}(x)}{\sum_{x \in s_{2,G}} M_G(x)} \quad \text{where } \hat{R}_{c,G}(x) = Y_{c,G}(x) - \hat{Y}_{c,G}(x) \quad (12c)$$

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. 13a. 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 $\hat{\mathbf{Z}}_{c,G}$. 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 $\bar{\hat{R}}_{2,G}$ in G . This correction insures that the *zero mean residual property* in F also holds within the small area G , which is per se not ensured by fitting the regression coefficients with data outside G .

$$\hat{Y}_{c,G,PSMALL} = \hat{\mathbf{Z}}_{c,G}^\top \hat{\beta}_{c,s_2} + \bar{\hat{R}}_{2,G} \quad (13a)$$

$$\begin{aligned} \hat{\mathbb{V}}(\hat{Y}_{c,G,PSMALL}) &= \hat{\mathbf{Z}}_{c,G}^\top \hat{\Sigma}_{\hat{\beta}_{c,s_2}} \hat{\mathbf{Z}}_{c,G} + \hat{\beta}_{c,s_2}^\top \hat{\Sigma}_{\hat{\mathbf{Z}}_{c,G}} \hat{\beta}_{c,s_2} \\ &\quad + \frac{1}{n_{2,G}(n_{2,G}-1)} \sum_{x \in s_{2,G}} \left(\frac{M_G(x)}{\bar{M}_{2,G}} \right)^2 (\hat{R}_{c,G}(x) - \bar{\hat{R}}_{2,G})^2 \end{aligned} \quad (13b)$$

The estimated design-based variance of $\hat{Y}_{c,G,PSMALL}$ is given by Eq. 13b. 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 ([Mandallaz et al, 2016](#), pg.14) which ensure calibration properties of the auxiliary variables on the terrestrial sample.

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

$$\hat{\Sigma}_{\hat{\mathbf{Z}}_{c,G}} = \frac{1}{n_{1,G}(n_{1,G}-1)} \sum_{x \in s_{1,G}} \left(\frac{M_G(x)}{\bar{M}_{1,G}} \right)^2 (\mathbf{Z}_{c,G}(x) - \hat{\mathbf{Z}}_{c,G})(\mathbf{Z}_{c,G}(x) - \hat{\mathbf{Z}}_{c,G})^\top \quad (14)$$

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

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. 15a) 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 $\hat{\mathbf{Z}}_{c,G}$. Note that the bias correction term using the empirical residuals (Eq. 13a) can no longer be applied. The PSYNTH estimator thus has a potential unobservable design-based bias.

$$\hat{Y}_{c,G,PSYNTH} = \hat{\mathbf{Z}}_{c,G}^\top \hat{\boldsymbol{\beta}}_{c,s_2} \quad (15a)$$

$$\hat{\mathbb{V}}(\hat{Y}_{c,G,PSYNTH}) = \hat{\mathbf{Z}}_{c,G}^\top \hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}_{c,s_2}} \hat{\mathbf{Z}}_{c,G} + \hat{\boldsymbol{\beta}}_{c,s_2}^\top \hat{\Sigma}_{\hat{\mathbf{Z}}_{c,G}} \hat{\boldsymbol{\beta}}_{c,s_2} \quad (15b)$$

The uncertainty of the model predictions can also no longer be considered in the variance estimation (Eq. 15b). The synthetic estimator will therefore usually have a smaller variance than estimators incorporating the model uncertainties, 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. 16) has been proposed by [Mandalaz \(2013a\)](#) 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. $\tilde{R}_c = \tilde{R}_{c,G} = 0$. 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 $\mathbf{Z}_c^\top(x) = (\mathbf{Z}_c^\top(x), I_{c,G}(x))$ and the new regression coefficient using $\mathbf{Z}_c(x)$ instead of $\mathbf{Z}_c(x)$ 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. 12. A decomposition of $\hat{\boldsymbol{\theta}}_{s_2}$ reveals that the residual correction term is now included in the regression coefficient $\hat{\boldsymbol{\theta}}_{s_2}$.

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

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

However, it is important to note that $\hat{R}_{c,G} = 0$ under the extended regression model only holds if the sample plots x_1, \dots, x_l of a cluster are *all* either inside or outside the small area, i.e. $M_G(x) \equiv M(x)$, and thus $I_{c,G}(x) = \frac{M_G(x)}{M(x)}$ can only take the values 1 or 0. [Mandalaz et al \(2016\)](#) 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 accuracies were quantified by the *estimation error*, which is the ratio of the standard error and the point estimate:

$$error(\hat{Y})[\%] = \frac{\sqrt{\hat{\mathbb{V}}(\hat{Y})}}{\hat{Y}} * 100 \quad (17)$$

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

$$CI_{1-\alpha}(\hat{Y}) = \left[\hat{Y} - t_{n_2-1,1-\frac{\alpha}{2}} \sqrt{\hat{\mathbb{V}}(\hat{Y})}, \hat{Y} + t_{n_2-1,1-\frac{\alpha}{2}} \sqrt{\hat{\mathbb{V}}(\hat{Y})} \right] \quad (18)$$

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

$$CI_{1-\alpha}(\hat{Y}) = \left[\hat{Y} - t_{n_{2,G}-1,1-\frac{\alpha}{2}} \sqrt{\hat{\mathbb{V}}(\hat{Y})}, \hat{Y} + t_{n_{2,G}-1,1-\frac{\alpha}{2}} \sqrt{\hat{\mathbb{V}}(\hat{Y})} \right] \quad (19)$$

For the PSYNTH estimates, the confidence intervals are

$$CI_{1-\alpha}(\hat{Y}) = \left[\hat{Y} - t_{n_2-p,1-\frac{\alpha}{2}} \sqrt{\hat{\mathbb{V}}(\hat{Y})}, \hat{Y} + t_{n_2-p,1-\frac{\alpha}{2}} \sqrt{\hat{\mathbb{V}}(\hat{Y})} \right] \quad (20)$$

In order to address the potential benefits of the small area estimators compared with the SRS approach, we calculated the *relative efficiency* (Eq. 21) which can be interpreted as the relative sample size under SRS needed to achieve the variance under the double sampling estimators.

$$rel.eff = \frac{\hat{\mathbb{V}}_{SRS}(\hat{Y})}{\hat{\mathbb{V}}_{SAE_{2phase}}(\hat{Y})} \quad (21)$$

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 km²) of the entire state area (19850 km²) covered by forest, RLP is one of the two states with the highest forest coverage among all federal states of Germany ([Thünen-Institut, 2014](#)). The forests of RLP are further characterized by a pronounced diversity in bioclimatic growing conditions that have strong influence on the local growth dynamics as well as tree species composition ([Gauer and Aldinger, 2005](#)) and are further characterized 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 ([LWaldG, 2000](#)). 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 small area (*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 were 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 cluster (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 frame of this survey, the plot center coordinates were re-measured with differential global satellite navigation system (DGPS) technique. Knowledge about the exact plot positions were considered crucial to provide optimal comparability between the terrestrially 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

5. CASE STUDY

RLP by Lamprecht et al (2017) 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 cluster on average and ranged between 11 and 64. Within the FR units, $n_{2,G}$ was considerably smaller with an average of 5 cluster and a range between 0 and 13.

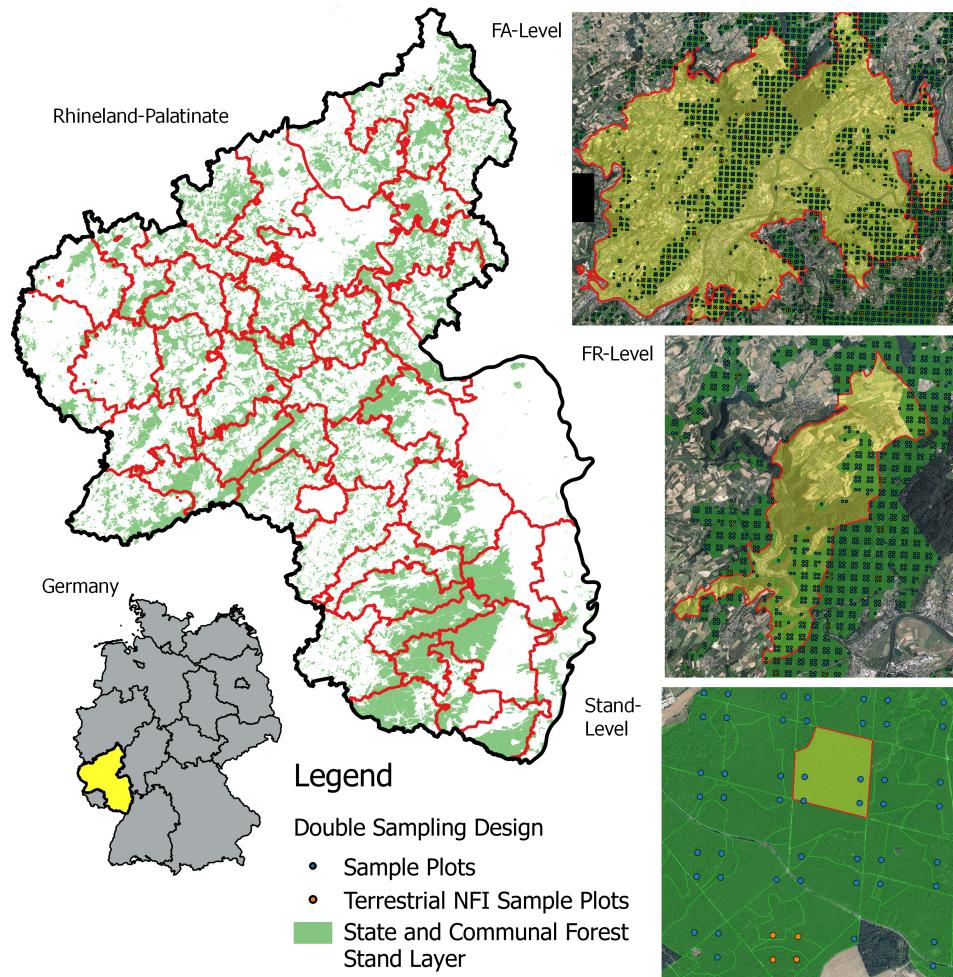


Figure 1: *Left:* Study Area with delineated FA forest management units. *Right:* Example for each of the three management units (from top to bottom): FA, FR and forest stand unit with overlayed the extended double sampling cluster design. *Green:* Forest stand polygon layer defining the forest area 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 kilometer grid to a grid size of 500x500 meters 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 cluster 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 cluster (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 cluster (range: 246 – 1022) in the FA units, and 88 cluster (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 was the identification of suitable auxiliary data available over the entire study area. Between the year 2003 and 2013, the topographic survey institution of RLP conducted an airborne laser scanning (ALS) acquisition over the entire federal state at leaf-off condition in order to derive a countrywide digital terrain and surface model. For this study, the recorded LiDAR 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 (Figure 2, top). The CHM (Fig. 2, top) was calculated as the difference between the digital terrain model (DTM) and the digital surface model (DSM) 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 \(2018\)](#). The CHM was considered to provide the most valuable information to be used in the OLS regression model for predicting the timber volume on sample plot and cluster level. However, the extended acquisition period of the ALS campaign led to substantial time gaps to the BWI3 survey of up to 10 years. In addition, the CHM exhibited severe quality variations in the CHM due to evolving ALS technology over the years. LiDAR acquisition recorded in 2002 and 2003 particularly showed a rather poor quality with about

only 0.04 point per m^2 , while more recently acquired 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 its non-available information thus 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 ([Stoffels et al, 2015](#)), 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 5 meters and was calculated from 22 bi-temporal satellite images (SPOT5 and RapidEye) by application of a spatially adaptive classification algorithm ([Stoffels et al, 2012](#)). As timber volume estimations on the tree level are 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 [White et al \(2016\)](#). 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 ([Kublin et al, 2013](#)) 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 (10 FR units respectively). 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 an area of 76 km^2 and 100 km^2 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. With respect to fitting the internal model for estimation, the tree species information was missing for 411 (7%) of the 5791 sample locations. 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 cluster

<i>Sampling frame</i>	$n_{1,plot}$	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

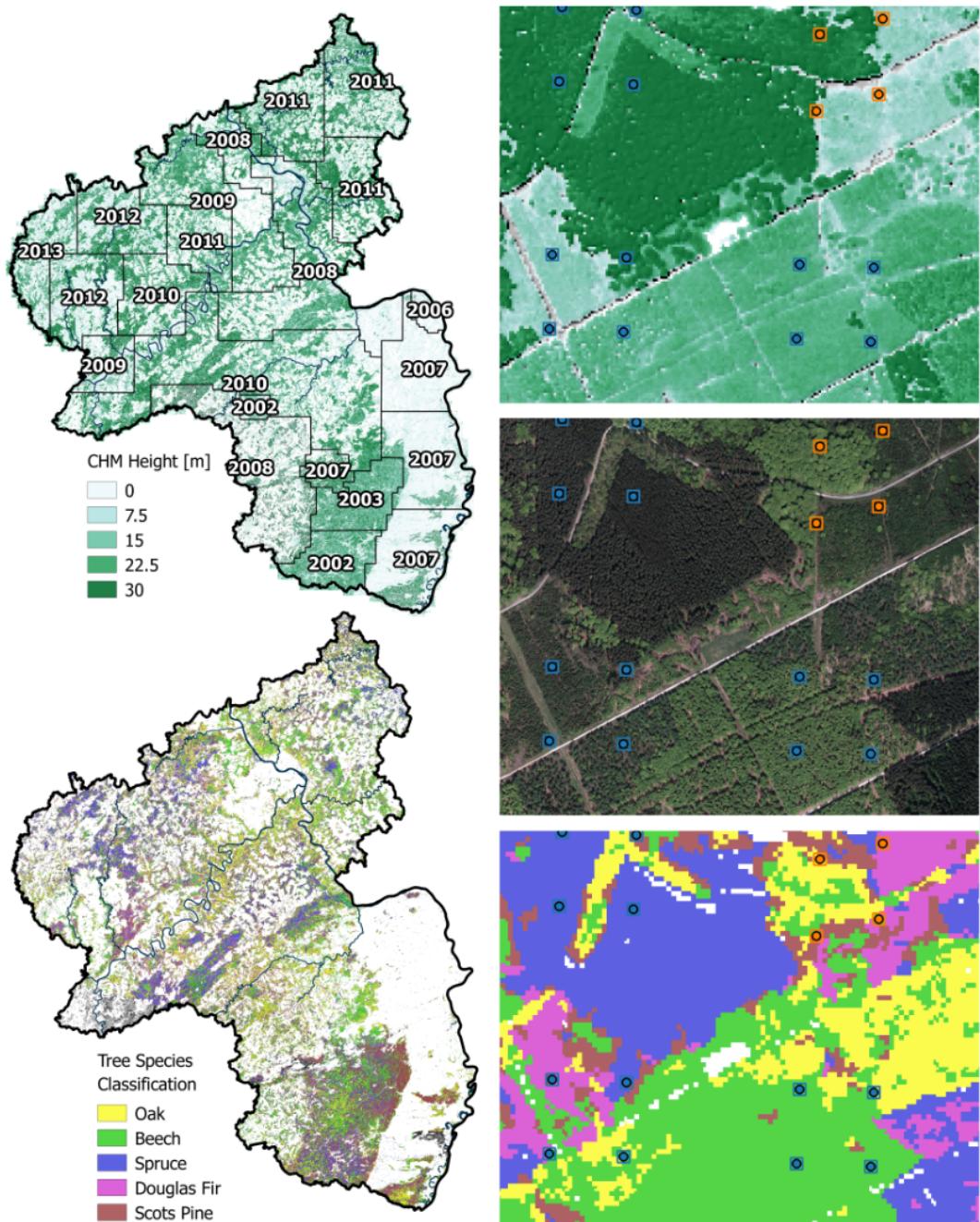


Figure 2: Left: CHM (top) and tree species classification map (bottom) available on the federal state level were used as auxiliary information. Right: Magnified illustration of the supports used to derive the explanatory variables from the auxiliary data.

5.5 Calculation of the explanatory variables

5.5.1 ALS canopy height model

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 predefined circle (*support*) with a radius of 12 meters. In order to correct for edge effects at the forest border, each support area was previously intersected with the state and communal forest area, which was defined by 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. Restricting the auxiliary data evaluation to the forest area was primarily a means to optimize the coherence between explanatory variables computed at the forest boundary and the corresponding local density, but also supports the consistency with the sampling frame (section 5.3). In particular, the BWI3 survey applies an edge correction at the forest border at the individual tree level by intersection of the trees' inclusion circle with the forest border. An adjustment for the fact that part of a trees' inclusion circle is outside the forest area is then realized by increasing its counting factor, which would otherwise lead to an underestimation of the local density. Since the ALS height values will usually drop to around zero outside the forest area, neglecting the edge correction of the support attenuates the mean canopy height value towards zero and thus increases the discrepancy to the local density.

We further derived the year of the ALS acquisition (*ALSyear*) as an additional categorical variable which was used to account for the time-lag between the CHM information and the terrestrial survey as well as to explain heterogeneity in the data introduced by the varying ALS quality. Adjustments were made to the original acquisition years by introducing an additional factor level *2008_I* for a subset of the 2008 acquisition where the quality turned out to be considerably poor due to a sensor error. In addition, the years 2006 and 2007 as well as 2012 and 2013 were pooled in order to increase the number of observations per factor level, resulting in nine categories in total (2002, 2003, 2007, 2008_I, 2008, 2009, 2010, 2011 and 2012).

5.5.2 Tree species map

The tree species map information was used to predict the main tree species of the sample trees at each sample plot (*treespecies*) as an additional categorical variable. This implied 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 particularly constitute parameters to be optimized in order

to achieve an optimal variance decomposition of the data that subsequently leads to the best possible model accuracy. A detailed analysis and description of the optimal processing parameters for the explanatory variables of the present data set is provided in [Hill et al \(2018\)](#). In this study, the authors also applied a calibration model to the initially derived *treespecies* variable that successfully removed the effects of misclassification errors on the regression model coefficients and increased the model accuracy. The calibration model consists of a decision tree from a random forest algorithm [Breiman \(2001\)](#) 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 $\sqrt{p} \approx 3$ of the predictors for each split. In a *second* step, we thus applied this calibration model to the *treespecies* variable derived at all sample locations s_1 . Table 3 gives the classification accuracies of the *treespecies* variable before and 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[%]		nref	nclass	
	uncalib	calib	uncalib	calib		uncalib	calib
Beech	29.56	22.31	46.11	47.02	883	566	419
Douglas Fir	33.04	24.78	47.50	48.72	230	160	117
Oak	47.75	11.07	24.34	48.48	289	567	66
Spruce	43.63	53.15	60.81	61.13	651	467	566
Scots Pine	55.87	22.91	29.59	46.07	179	338	89
Mixed	66.40	84.49	63.69	64.53	3152	3286	4127
Overall accuracy: 54.83% (61.96%)					5384	5384	

5.6 Regression Model

The model selection constituted a major part of the current study, especially because sophisticated challenges such as a) the heterogeneity in the remote sensing data, b) the identification of the optimal support sizes under angle count sampling and c) the incorporation of tree species information had to be addressed and investigated in order to identify the most accurate regression model realizable under the given data. We will here thus only provide a shortened summary of the extensive analysis carried out and refer to the study of [Hill et al \(2018\)](#) if the reader is interested in more detailed information on the subjects.

The model with highest adjusted R^2 and lowest RMSE was achieved with the variables *meanheight*, *meanheight*², *stddev* and *ALSyear* derived from the ALS data as well as the *treespecies* variable derived from the classification map as main effect terms. Additionally, interaction terms between *meanheight* and *ALSyear*, *stddev* and *ALSyear*, *meanheight* and *stddev*, and *meanheight* and *treespecies* were included. The model yielded an adjusted R^2 of

5. CASE STUDY

0.48 and an RMSE of 140.62 m³/ha (46.7%) on the plot level (table 4, full model). While the study of Hill et al (2018) only provide the model evaluations on the plot level, the two-phase estimators described in section 4.2 derive and apply the regression coefficients and the residuals on the aggregated cluster level. We thus re-evaluated the model as used in the estimators also on the cluster level and found the model accuracies to be higher than on the plot level (adjusted R^2 of 0.53 and RMSE of 101.61 m³/ha and 33.6%). A stratification to the ALS acquisition years substantially improved the model accuracy and thus proved to be an effective means in accounting for the artificially introduced noise in the data set due to ALS quality variations and time-gaps between the ALS and the terrestrial survey. However, the stratification led to highly unbalanced datasets when stratifying according to the *treespecies* variable. For this reason, a species individual modelling within each *ALSpyear* stratum remained infeasible, but might have further improved the model accuracy. An additional evaluation of the model residuals within each ALS acquisition year stratum (table 5) revealed that the model accuracies substantially varied between the strata (table 5). adjusted R^2 values above the average were particularly achieved in ALS acquisition years close to the terrestrial survey date (0.59 even up to 0.66 on the cluster level). This effect was substantially diminished when dropping the *ALSpyear* variable from the model term.

As depicted 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 model accuracy of the full and reduced model were found to be very similar on both the plot and cluster level. We thus assumed that the application of the reduced model would not cause substantially larger estimation errors as compared to the full model and subsequently performed a joint evaluation of the estimation results in section 6.

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

model terms	model	R^2_{adj}	RMSE	RMSE%
meanheight + stddev + meanheight ² +	full model	0.48	139.22	45.98
treespecies + ALSyear +		(0.53)	(101.61)	(33.57)
meanheight:treespecies +				
meanheight:ALSpyear + meanheight:stddev +				
stddev:ALSpyear				
meanheight + stddev + meanheight ² +	reduced model	0.45	144.13	47.60
ALSpyear + meanheight:ALSpyear +		(0.49)	(106.08)	(35.40)
meanheight:stddev + stddev:ALSpyear				

It should be mentioned that there is a major difference between the external and internal model approach concerning the treatment of outliers or leverage points for model fitting. In order to ensure the zero-mean-residual property of the PSMALL and EXTPSYNTH estimator under the internal model approach (section 4.2.1 and 4.2.3), all observations of the terrestrial

sample locations $x \in s_2$ used for estimation must also be included in the modelling frame. Excluding an observation from the model fit thus requires deleting it also from the sampling frame, which has to be regarded as an interference with the random sampling process the design-based inference relies on. For this reason, a removal of potential outliers or leverage points is, strictly speaking, only justified if the terrestrial response value or the explanatory variable value turns out to be truly erroneous (e.g. typos or erroneous measurements). If this is not the case, the removal of outliers or influential data points might increase the model fit, but to the cost of possible bias for the estimates (bias-variance trade-off).

We conducted a leverage and outlier detection on the plot level for the full regression model applying the usual criteria of *Leverages* and *Cook's Distance* as amongst others described in Fahrmeir et al (2013). The Cook's Distance values D_i did not exceed a value of 0.019, and thus were far apart from the often cited critical threshold of $D_i > 0.5$. However, the critical threshold of $2p/n$ for the leverage values (i.e. twice the average of the hat matrix' diagonal entries) was exceeded by 10% of the observations.

We conducted an Influence Analysis, i.e. leverage / outlier detection, for the 'final model'. We here considered the usually applied criteria of *Leverages* and *Cook's Distance* as amongst others described in Fahrmeir 2013 {Regression: models, methods and applications. Springer Science & Business Media. page 164-167}. The critical threshold of $2p/n$ (i.e. twice the average of the hat matrix' diagonal entries), was exceeded by 10% of the observations. However, only 3% of these leverage points were assigned to studentized residuals ζ_1 or ζ_{-1} . Leaving these 3% of points out and recalculating the final model lead to a R^2_{adj} of 0.494 compared to 0.485 when including them in the regression. The Cook's Distance values D_i did not exceed a value of 0.019, and thus were far apart from the often cited critical threshold of $D_i > 0.5$. Based on these findings, we decided not to remove observations from the modelling data set. We added a paragraph commenting on this issue of influential data points in section 3.3 'Final Regression Model', page 9 RHS line 690-707.

5. CASE STUDY

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

$ALSyear$	$Area_{ALSyear}$	R^2	RMSE	RMSE%	n
2012	2807	0.61 (0.66)	135.84 (101.99)	44.87 (30.61)	408 (156)
2011	4361	0.57 (0.59)	146.21 (109.78)	48.29 (34.23)	883 (354)
2010	4182	0.51 (0.58)	120.90 (86.88)	39.93 (31.44)	1171 (420)
2009	2100	0.42 (0.45)	133.42 (102.07)	44.07 (36.63)	559 (218)
2008	2968	0.48 (0.53)	130.38 (95.16)	43.06 (35.53)	701 (247)
2008_1	2116	0.33 (0.33)	175.43 (135.90)	57.94 (38.43)	394 (157)
2007	3498	0.46 (0.48)	136.47 (94.42)	45.08 (30.41)	418 (135]
2003	602	0.27 (0.29)	154.48 (98.96)	51.02 (30.57)	529 (145)
2002	775	0.44 (0.61)	141.55 (85.4)	46.75 (26.76)	314 (97)

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 substantial 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			Errors [%]		
		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 errors

On both small area levels, the design-unbiased estimators PSMALL and EXTPSYNTH led to a substantial reduction in the estimation errors 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 the EXTPSYNTH and PSMALL estimator (table 6). The cumulative error distribution (fig. 3, left) reveals that under the SRS estimator, errors $\leq 5\%$ were achieved for 17% of the FA units (8 of 45). This proportion could be increased to 62% (28 FA units) and

6. RESULTS

73% (33 FA units) by application of the PSMALL and EXTPSYNTH estimator. 95% of all estimates exhibited errors $\leq 9.5\%$ under the SRS estimator and $\leq 6.6\%$ when using PSMALL or EXTPSYNTH. Estimation errors higher than 10% only appeared twice for each of the three estimators.

The error reduction by the PSMALL and EXTPSYNTH estimator was even more pronounced on the FR level (fig. 3, right), although the overall error niveau was substantially higher than on FA level. The average error under the SRS estimator was 18.3%, while it was 11.3% and 12.2% under the PSMALL and EXTPSYNTH estimator (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 $\leq 20\%$. In comparison, the SRS estimates resulted in errors $\leq 36.6\%$ for 95% of the 388 FR units.

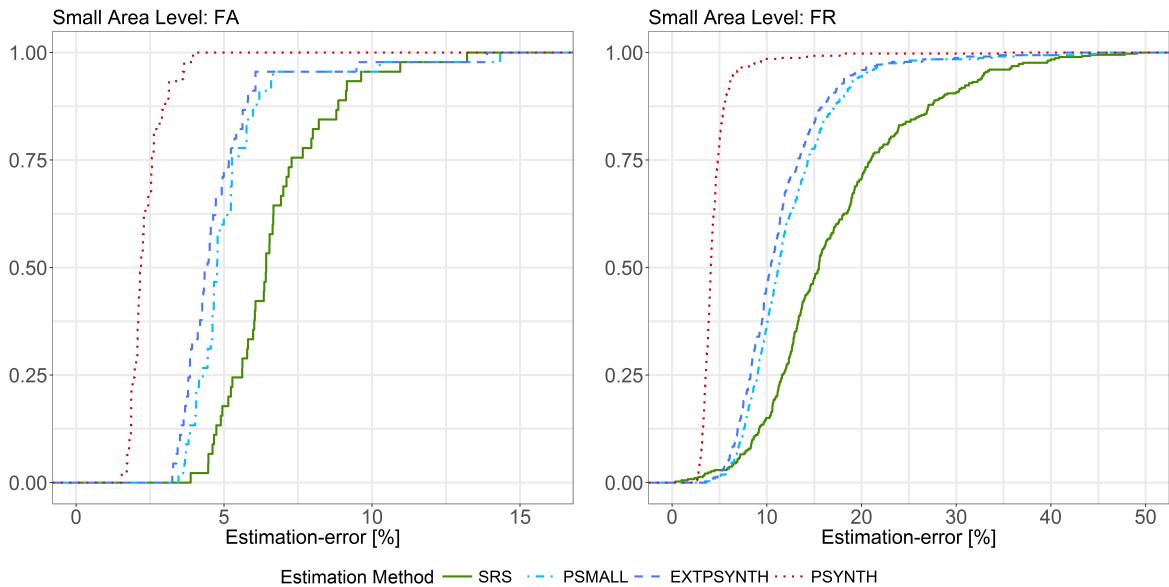


Figure 3: Cumulative distribution of estimation errors under the simple random sampling (SRS), the pseudo small (PSMALL), the extended pseudo synthetic (EXTPSYNTH) and the pseudo synthetic (PSYNTH) estimator. *Left:* Results for the 45 FA units. *Right:* Results for the 388 (SRS), 321 (PSMALL / EXTPSYNTH) 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 such differences was the effect of one or more cluster not entirely being included in a small area unit, as this would constitute a violation of the EXTPSYNTH estimator. This violation was actually observed in 155 of the 321 FR units (48%). However, the affected FR units (depicted in red diamonds, fig. 4) did not show a significant divergence from the PSMALL variances with respect to the remaining unaffected FR units. The variance differences between the two estimators were thus due to the mathematical formulations of the PSMALL and EXTPSYNTH estimator, which are asymptotically equivalent only under large terrestrial sample sizes $n_{2,G}$ within the small area (Mandallaz et al, 2016, pp.17–18). An additional comparison of the absolute differences in the g-weight variance (fig. 4, right) revealed that large divergences did in fact particularly occur for small area units with small terrestrial sample sizes ($n_{2,G} \leq 5$). The differences decreased with increasing sample size and thus confirmed the asymptotic relationship between the two estimators. However, a comparison of the confidence intervals of PSMALL and EXTPSYNTH revealed that the variance differences did not lead to statistically significant point estimates. The results indicate that the adjustment of the intercept term to the data within a small area used in the EXTPSYNTH estimator will in general yield slightly smaller variances than the PSMALL estimator.

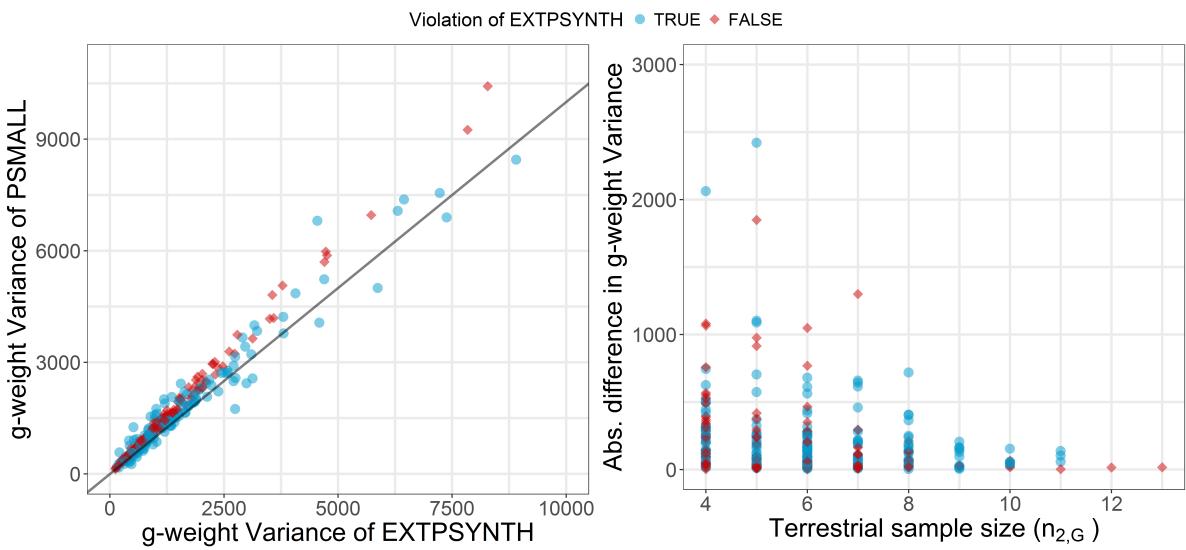


Figure 4: *Left:* Comparison of the g-weight variance between the PSMALL and the EXTPSYNTH estimator for the 321 FR units. *Right:* Difference in g-weight variance between the PSMALL and the EXTPSYNTH estimator in dependence of the terrestrial data (n_{2G}) in the FR unit.

6.4 Reduction of SRS variance by PSMALL and EXTPSYNTH

A direct comparison of the achieved variances within the small area units revealed that the application of the design-unbiased estimators (PSMALL and EXTPSYNTH) led to a reduction of the SRS variance in all FA units. In 75% of the FA units, the EXTPSYNTH estimator was able to reduce the SRS variance by up to 54.1% (fig. 5). The reduction in variance was also 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 and the relative efficiencies reached even higher values (table 7). The PSMALL estimator again yielded slightly lower variance reductions and relative efficiencies due to the smaller variances produced by the EXTPSYNTH estimator (section 6.3). However, cases also occurred on the FR level where one or both two-phase estimators produced larger variance values than under the SRS estimator. This particularly happened in 19% (61) of the FR units under the EXTPSYNTH, and in 24% (76) of the FR units under the PSMALL estimator.

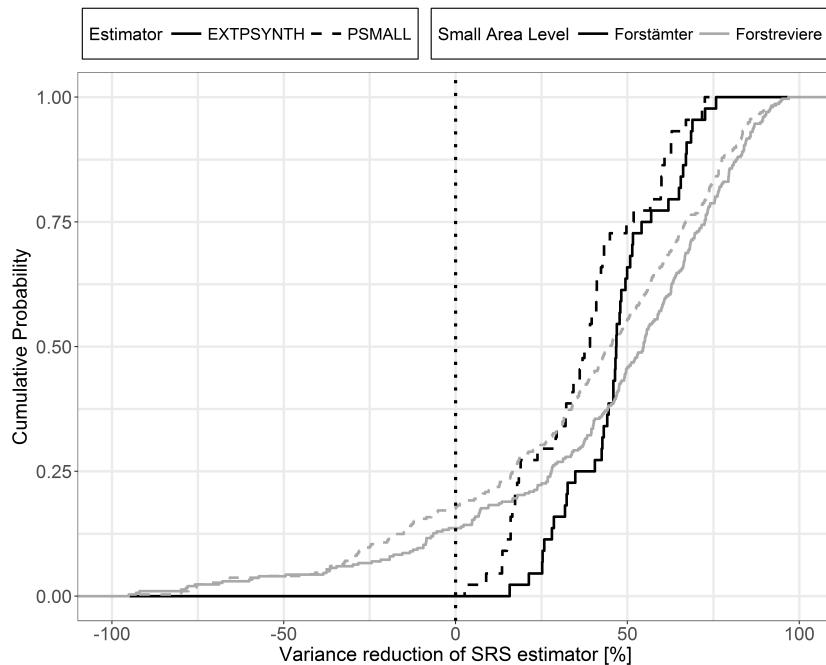


Figure 5: Cumulative distribution of variance reduction by the PSMALL and EXTPSYNTH compared to the SRS estimator for the 45 FA and 321 FR units.

One possible reason for an increased variance by the two-phase estimators was supposed to be 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. 13b), 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 that expresses the model performance within the small area (term 3). Figure 6 illustrates the reduction or increase of the SRS variance by PSMALL for all FR units in dependence of the residual variance term proportion on the g-weight variance. Two issues

6. RESULTS

become obvious: firstly, the residual term generally constitutes the dominating part of the g-weight variance for all PSMALL estimates (around 84% on average). FR units for which the PSMALL estimator produced larger variances than the SRS estimator did not differ distinctly from the pattern of the remaining FR units. The same was also found when looking at the absolute residual term values (fig. 7). Secondly, FR units with the highest variance increases compared to SRS were particularly observed under small terrestrial sample sizes $n_{2,G} = 4$. The average increase in variance for $n_{2,G} = 4$ was 272% compared to 62% for $n_{2,G} > 4$. In comparison, the average reduction in SRS variances was around 50% for both sampling sizes $n_{2,G} = 4$ and $n_{2,G} > 4$. These findings suggest that moderate or poor model fits within a small area can cause the PSMALL and EXTPSYNTH estimator to produce higher variances than the SRS estimator particularly under small terrestrial sample sizes. However, small terrestrial sample sizes can also have the effect that the SRS estimator might 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 evidence for this hypothesis, a definite proof is practically infeasible.

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

District level	Estimator	Reduction of SRS variance [%]			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

6. RESULTS

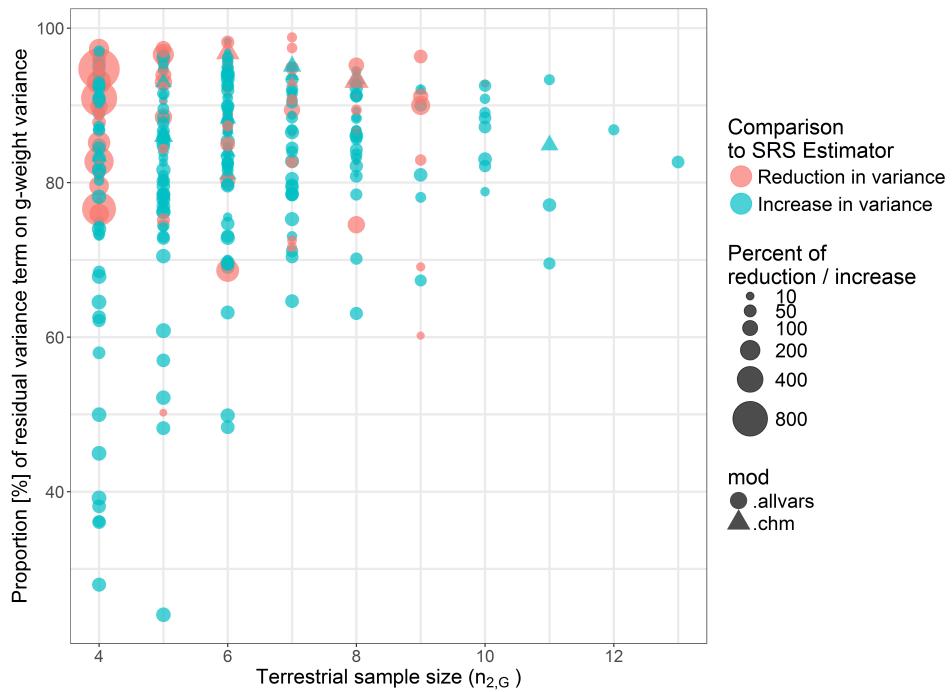


Figure 6

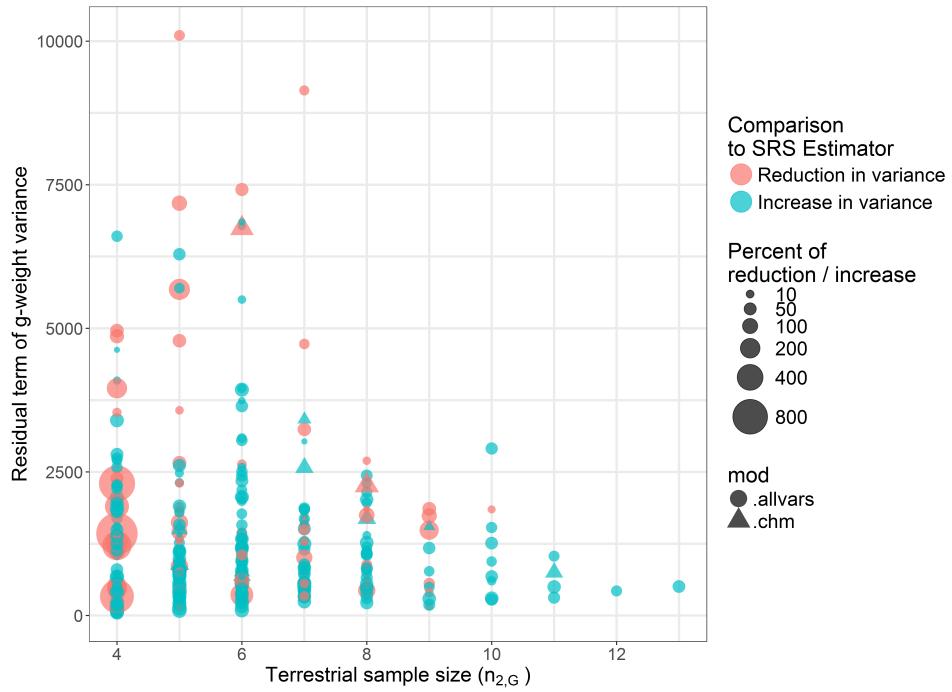


Figure 7

References

- Bitterlich W (1984) The relascope idea. Relative measurements in forestry. Commonwealth Agricultural Bureaux
- Böckmann T, Saborowski J, Dahm S, Nagel J, Spellmann H (1998) A new conception for forest inventory in lower saxony. *Forst und Holz* (Germany)
- Breidenbach J, Astrup R (2012) Small area estimation of forest attributes in the norwegian national forest inventory. *European Journal of Forest Research* 131(4):1255–1267, DOI 10.1007/s10342-012-0596-7
- Breiman L (2001) Random forests. *Machine learning* 45(1):5–32, DOI 10.1023/A:1010933404324
- Bundesministerium für Ernährung LuV (2011) Aufnahmeanweisung für die dritte Bundeswaldinventur BWI3 (2011 - 2012). URL <https://www.bundeswaldinventur.de/index.php?id=421>
- Fahrmeir L, Kneib T, Lang S, Marx B (2013) Regression: models, methods and applications. Springer Science & Business Media, DOI 10.1007/978-3-642-34333-9, URL <http://www.springer.com/us/book/9783642343322>
- Gauer J, Aldinger E (2005) Waldökologische naturräume deutschlands-wuchsgebiete. Mitteilungen des Vereins für Forstliche Standortskunde und Forstpflanzenzüchtung 43:281–288
- Goerndt ME, Monleon VJ, Temesgen H (2011) A comparison of small-area estimation techniques to estimate selected stand attributes using lidar-derived auxiliary variables. *Canadian journal of forest research* 41(6):1189–1201
- Grafström A, Schnell S, Saarela S, Hubbell S, Condit R (2017) The continuous population approach to forest inventories and use of information in the design. *Environmetrics*
- Gregoire TG, Valentine HT (2007) Sampling strategies for natural resources and the environment. CRC Press
- Hill A, Massey A (2017) forestinventory: Design-Based Global and Small-Area Estimations for Multiphase Forest Inventories. R package version 0.3.1. URL <https://cran.r-project.org/web/packages/forestinventory/index.html>
- Hill A, Buddenbaum H, Mandallaz D (2018) 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, submitted for publication
- Koch B (2010) 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* 65(6):581–590, DOI 10.1016/j.isprsjprs.2010.09.001, URL <https://doi.org/10.1016/j.isprsjprs.2010.09.001>

- Köhl M, Magnussen SS, Marchetti M (2006) Sampling methods, remote sensing and GIS multiresource forest inventory. Springer Science & Business Media
- Kublin E, Breidenbach J, Kändler G (2013) A flexible stem taper and volume prediction method based on mixed-effects b-spline regression. European journal of forest research 132(5-6):983–997, DOI 10.1007/s10342-013-0715-0
- Kuliešis A, Tomter SM, Vidal C, Lanz A (2016) 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 73(4):857–869
- Lamprecht S, Hill A, Stoffels J, Udelhoven T (2017) A machine learning method for co-registration and individual tree matching of forest inventory and airborne laser scanning data. Remote Sensing 9(5), DOI 10.3390/rs9050505, URL <http://www.mdpi.com/2072-4292/9/5/505>
- von Lüpke N (2013) Approaches for the optimisation of double sampling for stratification in repeated forest inventories. PhD thesis, University of Göttingen
- von Lüpke N, Hansen J, Saborowski J (2012) A three-phase sampling procedure for continuous forest inventory with partial re-measurement and updating of terrestrial sample plots. European Journal of Forest Research 131(6):1979–1990, DOI 10.1007/s10342-012-0648-z
- WLWaldG (2000) Landeswaldgesetz Rheinland-Pfalz (Forest Act Rhineland-Palatinate). URL <http://www.landesrecht.rlp.de>, Rhineland-Palatinate, Germany
- Magnussen S, Mandallaz D, Breidenbach J, Lanz A, Ginzler C (2014) National forest inventories in the service of small area estimation of stem volume. Canadian Journal of Forest Research 44(9):1079–1090
- Magnussen S, Mauro F, Breidenbach J, Lanz A, Kändler G (2017) Area-level analysis of forest inventory variables. European Journal of Forest Research pp 1–17
- Mandallaz D (2008) Sampling techniques for forest inventories. CRC Press, DOI 10.1201/9781584889779, URL <https://doi.org/10.1201/9781584889779>
- Mandallaz D (2013a) Design-based properties of some small-area estimators in forest inventory with two-phase sampling. Canadian Journal of Forest Research 43(5):441–449, DOI 10.1139/cjfr-2012-0381, URL <https://doi.org/10.1139/cjfr-2012-0381>
- Mandallaz D (2013b) A three-phase sampling extension of the generalized regression estimator with partially exhaustive information. Canadian Journal of Forest Research 44(4):383–388, DOI 10.1139/cjfr-2013-0449, URL <https://doi.org/10.1139/cjfr-2013-0449>
- Mandallaz D, Breschan J, Hill A (2013) New regression estimators in forest inventories with two-phase sampling and partially exhaustive information: a design-based monte carlo approach with applications to small-area estimation. Canadian Journal of Forest Research 43(11):1023–1031, DOI 10.1139/cjfr-2013-0181, URL <https://doi.org/10.1139/cjfr-2013-0181>

- Mandallaz D, Hill A, Massey A (2016) Design-based properties of some small-area estimators in forest inventory with two-phase sampling - revised version. Tech. rep., Department of Environmental Systems Science, ETH Zurich, DOI 10.3929/ethz-a-010579388, URL <https://doi.org/10.3929/ethz-a-010579388>
- Massey A, Mandallaz D, Lanz A (2014) 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 44(10):1177–1186, DOI 10.1139/cjfr-2014-0152, URL <https://doi.org/10.1139/cjfr-2014-0152>
- Naesset E (2014) Area-based inventory in norway – from innovation to an operational reality. In: Forest Applications of Airborne Laser Scanning - Concepts and Case Studies, Springer, chap 11, pp 216–240, DOI 10.1007/978-94-017-8663-8
- Polley H, Schmitz F, Hennig P, Kroher F (2010) Germany. In: National Forest Inventories - Pathways for Common Reporting, Springer, chap 13, pp 223–243, DOI 10.1007/978-90-481-3233-1
- Rao JN (2015) Small-Area Estimation. Wiley Online Library
- Saborowski J, Marx A, Nagel J, Böckmann T (2010) Double sampling for stratification in periodic inventories—finite population approach. Forest ecology and management 260(10):1886–1895
- Särndal CE, Swensson B, Wretman J (2003) Model assisted survey sampling. Springer Science & Business Media
- Schmitz F, Polley H, Hennig P, Dunger K, Schwitzgebel F (2008) Die zweite Bundeswaldinventur - BWI2: Inventur- und Auswertmethoden, Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz (Hrsg)
- Steinmann K, Mandallaz D, Ginzler C, Lanz A (2013) 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 28(4):373–385
- Stoffels J, Mader S, Hill J, Werner W, Ontrup G (2012) Satellite-based stand-wise forest cover type mapping using a spatially adaptive classification approach. European journal of forest research 131(4):1071–1089
- Stoffels J, Hill J, Sachtleber T, Mader S, Buddenbaum H, Stern O, Langshausen J, Dietz J, Ontrup G (2015) Satellite-based derivation of high-resolution forest information layers for operational forest management. Forests 6(6):1982–2013, DOI 10.3390/f6061982, URL <http://www.mdpi.com/1999-4907/6/6/1982>
- Thünen-Institut (2014) Dritte Bundeswaldinventur 2012. URL <https://bwi.info>, accessed: 2017-02-03

- White JC, Coops NC, Wulder MA, Vastaranta M, Hilker T, Tompalski P (2016) Remote sensing technologies for enhancing forest inventories: A review. *Canadian Journal of Remote Sensing* 42(5):619–641, DOI 10.1080/07038992.2016.1207484, URL <http://dx.doi.org/10.1080/07038992.2016.1207484>