

1

Article Draft

2

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

3

4

5

6 Large scale application to the federal state of Rhineland-Palatinate

7

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

8

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

9

^{**}Forest Service Rhineland-Palatinate

¹⁰ Corresponding author[†]: andreas.hill@usys.ethz.ch
¹¹ Co-authors: daniel.mandallaz@env.ethz.ch, joachim.langshausen@wald-rlp.de
¹² Publication date: February 15, 2018

¹³ **Abstract**

¹⁴ The German National Forest Inventory consists of a systematic grid of permanent sam-
¹⁵ ple plots and provides a reliable evidence-based assessment of the state and the development
¹⁶ of Germany's forests on national and federal state level in a 10 year interval. However, the
¹⁷ data have yet been scarcely used for estimation on smaller management levels such as forest
¹⁸ districts due to insufficient sample sizes within the area of interests and the implied large es-
¹⁹timation errors. In this study, we present a double-sampling extension to the existing German
²⁰National Forest Inventory (NFI) that allows for the application of recently developed design-
²¹based small area regression estimators. We illustrate the implementation of the estimation
²²procedure and evaluate its potential by the example of timber volume estimation on two small
²³scale management levels (45 and 405 forest district units respectively) in the federal German
²⁴state of Rhineland-Palatinate. An airborne laserscanning (ALS) derived canopy height model
²⁵and a tree species classification map based on satellite data were used as auxiliary data in an
²⁶ordinary least square regression model to produce the timber volume predictions on the plot
²⁷level. The results support that the suggested double-sampling procedure can substantially in-
²⁸crease estimation precision on both management levels: the two-phase estimators were able to
²⁹reduce the variance of the SRS estimator by 43% and 25% on average for the two management
³⁰levels respectively.

³¹ **Keywords.** National forest inventory, small area estimation, double sampling for regression
³²within strata, cluster sampling, LiDAR canopy height model, tree species classification

33 Contents

34	1 Introduction	1
35	2 Terrestrial sampling design of the German NFI	3
36	3 Double sampling in the infinite population approach	3
37	4 Estimators	4
38	4.1 Design-based SRS estimator for cluster sampling	4
39	4.2 Design-based small area regression estimators for cluster sampling	6
40	4.2.1 Pseudo Small Area Estimator (PSMALL)	7
41	4.2.2 Pseudo Synthetic Estimator (PSYNTH)	8
42	4.2.3 Extended Pseudo Synthetic Estimator (EXTPSYNTH)	8
43	4.3 Measures of estimation accuracy	9
44	5 Case study	10
45	5.1 Study area and small area units	10
46	5.2 Terrestrial sample	10
47	5.3 Extension to double sampling design	11
48	5.4 Auxiliary data	12
49	5.4.1 LiDAR canopy height model	12
50	5.4.2 Tree species map	13
51	5.5 Calculation of the explanatory variables	15
52	5.5.1 ALS canopy height model	15
53	5.5.2 Tree species map	15
54	5.6 Regression Model	16
55	6 Results	19
56	6.1 General estimation results	19
57	6.2 Estimation errors	19
58	6.3 Comparison of PSMALL and EXTPSYNTH	21
59	6.4 Reduction of SRS variance by PSMALL and EXTPSYNTH	22
60	7 Discussion	24
61	7.1 Performance of estimators	24
62	7.2 Auxiliary data	25
63	8 Conclusion	26
64	A Appendix	27

65 List of Figures

66 1	Left: Study Area with delineated FA forest management units. Right: Exam-	
67 ple for each of the three management units (from top to bottom): FA, FR and		
68 forest stand unit with overlayed the extended double sampling cluster design.		
69 <i>Green</i> : Forest stand polygon layer defining the forest area of this study.	11	
70 2	Left: CHM (top) and tree species classification map (bottom) available on	
71 the federal state level were used as auxiliary information. Right: Magnified		
72 illustration of the supports used to derive the explanatory variables from the		
73 auxiliary data.	14	
74 3	Cumulative distribution of estimation errors under the simple random sam-	
75 pling (SRS), the pseudo small (PSMALL), the extended pseudo synthetic		
76 (EXTPSYNTH) and the pseudo synthetic (PSYNTH) estimator. Left: Re-		
77 sults for the 45 FA units. Right: Results for the 388 (SRS), 321 (PSMALL /		
78 EXTPSYNTH) and 403 (PSYNTH) FR units.	20	
79 4	Left: Comparison of the g-weight variance between the PSMALL and the	
80 EXTPSYNTH estimator for the 321 FR units. Right: Difference in g-weight		
81 variance between the PSMALL and the EXTPSYNTH estimator in depen-		
82 dence of the terrestrial data ($n2G$) in the FR unit.	21	
83 5	Cumulative distribution of variance reduction by the PSMALL and EXTP-	
84 SYNTH compared to the SRS estimator for the 45 FA and 321 FR units. . . .	22	
85 6	Increase or reduction of the SRS variance by application of the PSMALL esti-	
86 mator for all 321 FR units in dependence of a) the terrestrial sample size in the		
87 FR units (x -axis) and b) the residual term proportion of the PSMALL g-weight		
88 variance (y -axis).	23	

89 **List of Tables**

90 1	Descriptive statistics of the forest observed on NFI sample plots located within 91 communal and state forest area (n=5791).	12
92 2	Sample size for each phase in entire study area. $n_{\{1,2\},plots}$: number of plots. 93 $n_{\{1,2\}}$: number of cluster	13
94 3	Classification accuracies of the <i>treespecies</i> variable before and after calibra- 95 tion. n_{ref} : number of terrestrial reference plots. n_{class} : number of classified 96 plots.	16
97 4	Accuracy metrics for the two OLS regression models on the cluster level. In- 98 teraction terms are indicated by ':'. () give the respective values on the plot 99 level.	17
100 5	R^2 , RMSE and RMSE% on the cluster level of the full regression model within 101 ALS acquisition year strata (<i>ALSpyear</i>). $Area_{ALSpyear}$: Area covered by ALS 102 acquisition given in km ² . n : number of validation data. () give the respective 103 values on the plot level.	18
104 6	Descriptive summary of point estimates and estimation errors on the two forest 105 district levels. N_u : number of evaluated small area units.	19
106 7	Descriptive summary of SRS variance reduction and relative efficiencies on 107 the two forest district levels. N_u : number of evaluated small area units.	22

108 **1 Introduction**

109 The German National Forest Inventory (NFI) provides reliable evidence-based and accu-
110 rate information of the current state and the development of Germanys forest over time. The
111 NFI thereby has to satisfy various information needs and amongst others reports to public and
112 state forestry administrations, wood-based industries and the public on the national level, as
113 well as to the Food and Agriculture Organization of the United Nations (FAO) and United Na-
114 tions Framework Convention on Climate Change (UNFCCC) on the international level ([Polley](#)
115 [et al, 2010](#)). At the current time, the inventory design of the German NFI solely rests upon
116 a terrestrial cluster inventory that is carried out at sample locations systematically distributed
117 over the entire forest state area of Germany. As this implies covering a large area of 114'191
118 ha ([Thünen-Institut, 2014](#)), the sample size has been chosen according to satisfy high esti-
119 mation accuracies for forest attributes on the national and federal state level. This however
120 leads to very low sampling intensities and consequently, sample sizes often drop dramatically
121 when entering spatial units below the federal state level. This is particularly true for forest
122 management levels such as forest districts for which the estimation uncertainties turn out be
123 unacceptably large due to the very limited number of sample plots within these units. For this
124 reason, the German NFI data have not yet been extensively incorporated in operational forest
125 planning on forest district management levels. In most German federal states, management
126 strategies are thus still based on expert judgements from time-consuming standwise invento-
127 ries (SFI), which are prone to systematic deviations [Kuliešis et al \(2016\)](#) and do not provide
128 any measure of uncertainty.

129 Some German federal states, such as Lower Saxony, have approached this problem by es-
130 tablishing a regional Forest District Inventory (FDI) with a much higher sampling density
131 than used by the NFI in order to base their regional management strategies on quantitative
132 and accurate information ([Böckmann et al, 1998](#)). However, such FDIs are cost-intensive and,
133 facing increasing restrictions in budget and staff resources, there has been a need for more
134 cost-efficient inventory methods ([von Lüpke, 2013](#)). One method which has proven to be ef-
135 ficient is double-or two-phase sampling ([Särndal et al, 2003; Gregoire and Valentine, 2007;](#)
136 [Köhl et al, 2006; Mandallaz, 2008](#)). Double sampling incorporates inexpensive auxiliary infor-
137 mation and can be used to either increase estimation precision under given terrestrial sample
138 size, or maintain estimation precision under reduced terrestrial sample size. A double sam-
139 pling for stratification procedure has e.g. been used in the FDI of Lower Saxony ([Saborowski](#)
140 [et al, 2010](#)), and [Grafström et al \(2017\)](#) lately illustrated how to use the auxiliary informa-
141 tion to determine optimised balanced terrestrial sample designs. Recent studies have lately
142 extended double-sampling to triple-sampling estimation methods using auxiliary information
143 in two different sampling intensities. An example can be found in [von Lüpke et al \(2012\)](#) who
144 illustrated an extension of the existing two-phase FDI Lower Saxony to a three-phase design
145 that uses updates of past inventory data as additional auxiliary information and allows for a
146 significant reduction of the terrestrial sample size in intermediate inventories. An other exam-
147 ple is [Massey et al \(2014\)](#) who developed a triple-sampling extension based on the ideas of
148 [Mandallaz \(2013b\)](#) for the Swiss NFI that can significantly reduce the increase in estimation
149 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-*

194 *proach*) and is also robust to heteroscedasticity of the model residuals. The results from this
195 study were considered to provide valuable information whether the suggested procedure might
196 be a cost-saving alternative to a regional FDI.

197 2 Terrestrial sampling design of the German NFI

198 The German National Forest Inventory (German NFI) is a periodic inventory that is carried
199 out every 10 years over the entire forest area of Germany. The most recent inventory (BWI3)
200 was conducted in the years 2011 and 2012. While information was originally gathered at a
201 systematic 4x4 km grid, some federal states such as Rhineland-Palatinate have switched to a
202 densified 2x2 km grid. The German NFI uses a cluster sampling design, which means that a
203 sample unit consists of maximal four sample locations (also referred to as *sample plots*) that
204 are arranged in a square (so called *cluster*) with a side length of 150 metres. The number of
205 plots per cluster can however vary between 1 and 4 depending on forest/non-forest decisions
206 by the field crews on the individual plot level ([Bundesministerium für Ernährung, 2011](#)). In
207 the field survey of the BWI3, sample trees for timber volume estimation are selected according
208 to the angle count sampling technique ([Bitterlich, 1984](#)), using a basal area factor (BAF) of 4
209 that is respectively adjusted for sample trees at the forest boundary by a geometric intersec-
210 tion of the boundary transect with the tree-individual inclusion circle ([Bundesministerium für](#)
211 [Ernährung, 2011](#)). A further inventory threshold for a tree to be recorded is a diameter at breast
212 height (dbh) of at least 7 cm. For each sample tree that is selected by this procedure, the dbh,
213 the absolute tree height, the tree diameter at 7 m (D7) and the tree species is measured and
214 used to calculate a volume estimation on the tree level. These volume estimations are based
215 on the application of tree species specific taper curves that are adjusted to the set of diameters
216 and corresponding height measurements taken from the respective sample tree ([Kublin et al,](#)
217 [2013](#)).

218 3 Double sampling in the infinite population approach

219 The estimators used in this study have been proposed by ([Mandallaz, 2013a; Mandallaz](#)
220 [et al, 2013](#)) and build upon the so called infinite population approach (IPA) in order to bridge
221 the inventory procedure and the derived information to the mathematics behind the estimators.
222 Therefore, we shall first provide a short introduction into this general estimation frame. We
223 start by assuming that the population P of trees $i \in 1, 2, \dots, N$ within a forest of interest F is ex-
224 actly defined, and each tree i has a directly or indirectly observable response variable Y_i (e.g. its
225 timber volume) that allows to specify the population mean Y (e.g., the average timber volume
226 per unit area) over F . If a full census of all tree population individuals is not possible, Y has to
227 be estimated by the conduction of an inventory. The infinite population approach assumes that
228 the spatial distribution of the local density $Y(x)$ (e.g., the timber volume per unit area) at each
229 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 [Mandalaz \(2008\)](#); [Mandalaz 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

4. ESTIMATORS

262 defined as the average of the individual sample plot densities $Y(x_l)$:

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

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

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

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

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

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

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

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

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

275 Using Eq. 5 in Eq. 3 yields the rewritten form of $Y(x_l)$ for angle count sampling that
 276 conforms to the definition used in the BWI3 algorithms ([Schmitz et al, 2008](#)):

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

277 where nha_i is the number of trees per hectare represented by tree i . The local densities on
 278 cluster level can then be used to derive the estimated spatial mean \hat{Y}_c and its estimated variance
 279 $\hat{\mathbb{V}}(\hat{Y}_c)$ for any given spatial unit for which $n_2 \geq 2$ (n_2 denoting the number of sample units, i.e.
 280 clusters):

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

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

281 4.2 Design-based small area regression estimators for cluster 282 sampling

283 All three considered small area estimators have in common that they use ordinary least
284 square (OLS) regression models to produce the predictions of the local density $Y_c(x)$ directly
285 on the cluster level c . We consider the *internal model approach*, where the vector of estimated
286 regression coefficients on the cluster level is found by "borrowing strength" from the entire
287 terrestrial sample s_2 of the current inventory:

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

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

288 $\mathbf{Z}_c(x)$ is the vector of explanatory variables on the cluster level, which is calculated as the
289 weighted average of the explanatory variables $\mathbf{Z}(x_l)$ on the individual plot levels x_1, \dots, x_l
290 (Eq.10). The weight $w(x_l)$ is the proportion of the support-area within the forest F used to
291 derive the explanatory variables from the raw auxiliary information.

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

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

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

295 with

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

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

298 In the following, we will give a short description of each small area estimator and refer to
299 Mandallaz (2013a); Mandallaz et al (2016, 2013) if the reader requires additional details or
300 proofs. The estimators have also been implemented in the R-package *forestinventory* (Hill
301 and Massey, 2017) which was used to compute all estimates in this study.

303

304 **4.2.1 Pseudo Small Area Estimator (PSMALL)**

305 All point information used for small area estimation is now restricted to that available at the
 306 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
 307 always based on the entire sample s_2 . We thus first define the following quantities on the small
 308 area level:

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

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

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

309 Note that the restriction to G , i.e. $I_G(x_l) = \{0, 1\}$, is made on the individual sample plot
 310 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
 311 small area. The asymptotically design-unbiased point estimate of *PSMALL* is then defined
 312 according to Eq. 14a. The first term estimates the small area population mean of G by applying
 313 the globally derived regression coefficients to the small area cluster means of the explanatory
 314 variables $\hat{\mathbf{Z}}_{c,G}$. The second term then corrects for a potential bias of the regression model
 315 predictions in the small area G by adding the mean of the empirical residuals $\bar{\hat{R}}_{2,G}$ in G . This
 316 correction insures that the *zero mean residual property* in F also holds within the small area
 317 G , which is per se not ensured by fitting the regression coefficients with data outside G .

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

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

318 The estimated design-based variance of $\hat{Y}_{c,G,PSMALL}$ is given by Eq. 14b. Basically, the
 319 first term constitutes the variance introduced by the uncertainty in the regression coefficients,
 320 whereas the second term expresses the variance caused by estimating the exact auxiliary mean
 321 in G using a non-exhaustive sample $s_{1,G}$. The third term is the variance of the model residuals
 322 and thus accounts for the inaccuracies of the model predictions. Note that the first term can
 323 also be rewritten using g-weights (Mandalaz et al, 2016, pg.14) which ensure calibration
 324 properties of the auxiliary variables on the terrestrial sample.

325
 326 The variance-covariance matrix of the auxiliary vector $\hat{\Sigma}_{\hat{\mathbf{Z}}_{c,G}}$ is thereby defined as

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

327 with $\bar{M}_{1,G} = \frac{\sum_{x \in s_{1,G}} M_G(x)}{n_{1,G}}$.
 328

329 **4.2.2 Pseudo Synthetic Estimator (PSYNTH)**

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

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

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

336 The uncertainty of the model predictions can also no longer be considered in the variance
 337 estimation (Eq. 16b). The synthetic estimator will therefore usually have a smaller variance
 338 than estimators incorporating the model uncertainties, but at the cost of a potential bias. Note
 339 that the PSYNTH estimator is still design-based, but one purely has to rely on the validity of
 340 the regression model within the small area as it is the case in the model-dependent framework.
 341

342 **4.2.3 Extended Pseudo Synthetic Estimator (EXTPSYNTH)**

343 The EXTPSYNTH estimator (Eq. 17) has been proposed by [Mandalaz \(2013a\)](#) as a trans-
 344 formed version of the PSMALL estimator that has the form of the PSYNTH estimator but
 345 remains asymptotically design unbiased. It has the advantage that the mean of the empirical
 346 model residuals of the OLS regression model for the entire area F and the small area G are
 347 by construction both zero at the same time, i.e. $\tilde{R}_c = \tilde{R}_{c,G} = 0$. This is realized by *extending*
 348 the auxiliary vector $\mathbf{Z}_c(x)$ by the indicator variable $I_{c,G}$ which takes the value 1 if the entire
 349 cluster lies within the small area G and 0 if the entire cluster is outside G , i.e. $I_{c,G}(x) = \frac{M_G(x)}{M(x)}$.
 350 The extended auxiliary vector thus becomes $\mathbf{Z}_c^\top(x) = (\mathbf{Z}_c^\top(x), I_{c,G}(x))$ and the new regression
 351 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
 352 are calculated by plugging in $\mathbf{Z}_c(x)$ in Eq. 13. A decomposition of $\hat{\boldsymbol{\theta}}_{s_2}$ reveals that the residual
 353 correction term is now included in the regression coefficient $\hat{\boldsymbol{\theta}}_{s_2}$.

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

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

354 However, it is important to note that $\tilde{R}_{c,G} = 0$ under the extended regression model only
 355 holds if the sample plots x_1, \dots, x_l of a cluster are *all* either inside or outside the small area,
 356 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 \(2016\)](#)
 357 assumed that the effects on the estimates should be negligible as the number of
 358 occasions where $M_G(x) < M(x)$ was considered to be small in practical implementations. It
 359 was thus a further objective of this study to investigate the actual occurrences and effects of
 360 this phenomenon by comparing the estimates of EXTPSYNTH to those of PSMALL.

361 **4.3 Measures of estimation accuracy**

362 The estimation accuracies were quantified by the *estimation error*, which is the ratio of the
 363 standard error and the point estimate:

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

364 We further calculated the 95% confidence interval for each estimate for visualization pur-
 365 pose. The confidence intervals can also be used for hypothesis testing whether the point esti-
 366 mates of the three estimators for a given small area are statistically different. The confidence
 367 intervals under SRS can be obtained as:

$$CI_{1-\alpha}(\hat{Y}) = \left[\hat{Y} - t_{n_2-1,1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y})}, \hat{Y} + t_{n_2-1,1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y})} \right] \quad (19)$$

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

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

369 For the PSYNTH estimates, the confidence intervals are

$$CI_{1-\alpha}(\hat{Y}) = \left[\hat{Y} - t_{n_2-p,1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y})}, \hat{Y} + t_{n_2-p,1-\frac{\alpha}{2}} \sqrt{\hat{V}(\hat{Y})} \right] \quad (21)$$

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

$$\text{rel. eff} = \frac{\hat{V}_{SRS}(\hat{Y})}{\hat{V}_{SAE_{2phase}}(\hat{Y})} \quad (22)$$

³⁷³ **5 Case study**

³⁷⁴ **5.1 Study area and small area units**

³⁷⁵ 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.

³⁹⁹ **5.2 Terrestrial sample**

⁴⁰⁰ 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

410 RLP by Lamprecht et al (2017) indicated that 80% of the plots should not exceed horizontal
411 DGPS errors of 8 meters. For 162 plots, the DGPS coordinates were replaced by their target
412 coordinates due to missingness or implausible values. The terrestrial sample size $n_{2,G}$ within
413 the FA units was 46 cluster on average and ranged between 11 and 64. Within the FR units,
414 $n_{2,G}$ was considerably smaller with an average of 5 cluster and a range between 0 and 13.

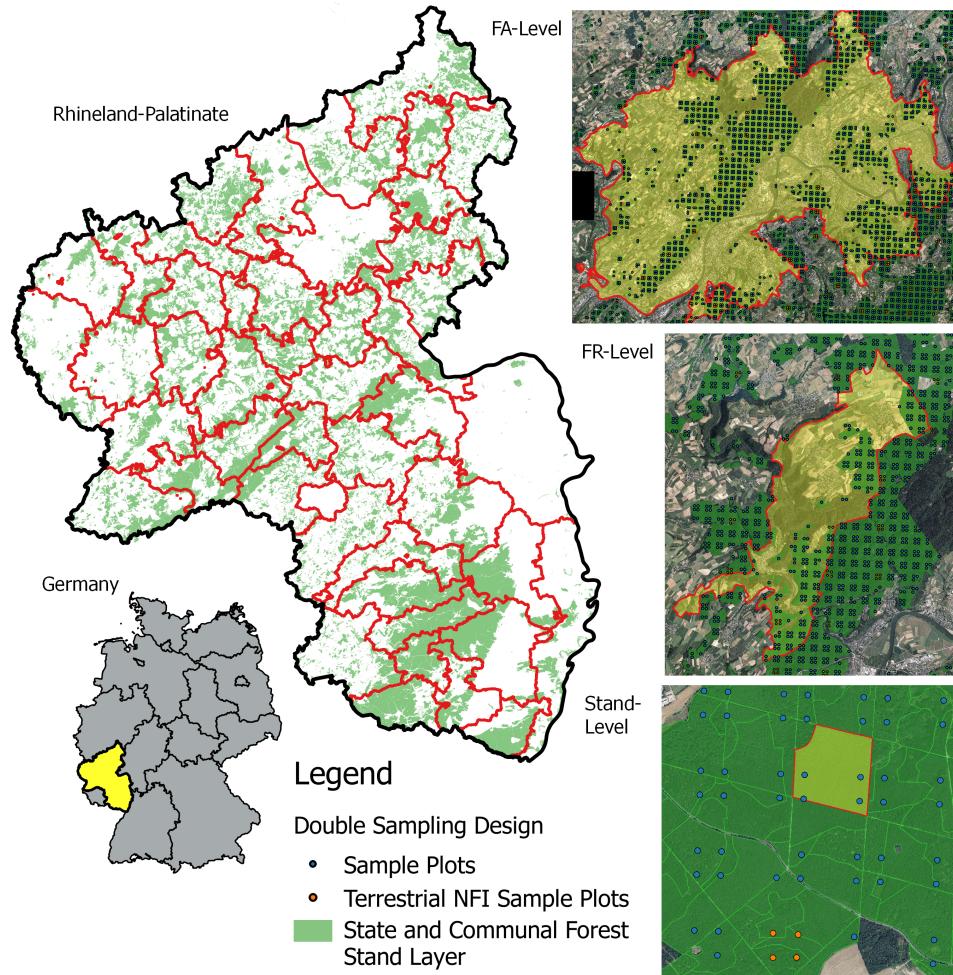


Figure 1: *Left:* Study Area with delineated FA forest management units. *Right:* Example for each of the three management units (from top to bottom): FA, FR and forest stand unit with overlayed the extended double sampling cluster design. *Green:* Forest stand polygon layer defining the forest area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

415 5.3 Extension to double sampling design

416 In order to apply the small area estimators (Section (4.2), the existing NFI design was
417 extended to a double sampling design by densifying the existing systematic 2x2 kilometer
418 grid to a grid size of 500x500 meters that constituted the large first phase s_1 in accordance to

5. CASE STUDY

419 Section 3 (Figure 1, right). The existing terrestrial phase s_2 was consequently integrated by
420 replacing the target coordinates of the respective s_1 cluster by the terrestrially measured DGPS
421 coordinates. For our study, we restricted the *sampling frame* to the communal and state forest.
422 The forest/non-forest decision for each plot was thereby made by a spatial intersection of the
423 plot center coordinates with a polygon layer of the communal and state forest stands provided
424 by the forest service. Using this stand layer provided the advantage to consistently apply the
425 same forest/non-forest definition to the entire sample s_1 in order to decide about excluding or
426 including a plot in the sampling frame. The terrestrial sample size n_2 was thus reduced to 2055
427 cluster (5791 plots). Table 1 provides a short descriptive summary about the volume densities
428 and the main attributes of the NFI plots located in the state and communal forest sampling
429 frame. The densification led to an average sample size $n_{1,G}$ of 759 cluster (range: 246 – 1022)
430 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

431 5.4 Auxiliary data

432 5.4.1 LiDAR canopy height model

433 A prerequisite for the application of the suggested two-phase small area estimators was the
434 identification of suitable auxiliary data available over the entire study area. Between the year
435 2003 and 2013, the topographic survey institution of RLP conducted an airborne laser scan-
436 ning (ALS) acquisition over the entire federal state at leaf-off condition in order to derive a
437 countrywide digital terrain and surface model. For this study, the recorded LiDAR data was
438 used to create a canopy height model (CHM) in raster format, providing discrete information
439 about the canopy surface height of the forest area in a spatial resolution of 5 meters (Figure 2,
440 top). The CHM (Fig. 2, top) was calculated as the difference between the digital terrain model
441 (DTM) and the digital surface model (DSM) that were derived by a Delauney interpolation
442 of the ground and first ALS pulses respectively. A more detailed description of the procedure
443 can be found in Hill et al (2018). The CHM was considered to provide the most valuable in-
444 formation to be used in the OLS regression model for predicting the timber volume on sample
445 plot and cluster level. However, the extended acquisition period of the ALS campaign led to
446 substantial time gaps to the BWI3 survey of up to 10 years. In addition, the CHM exhibited
447 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

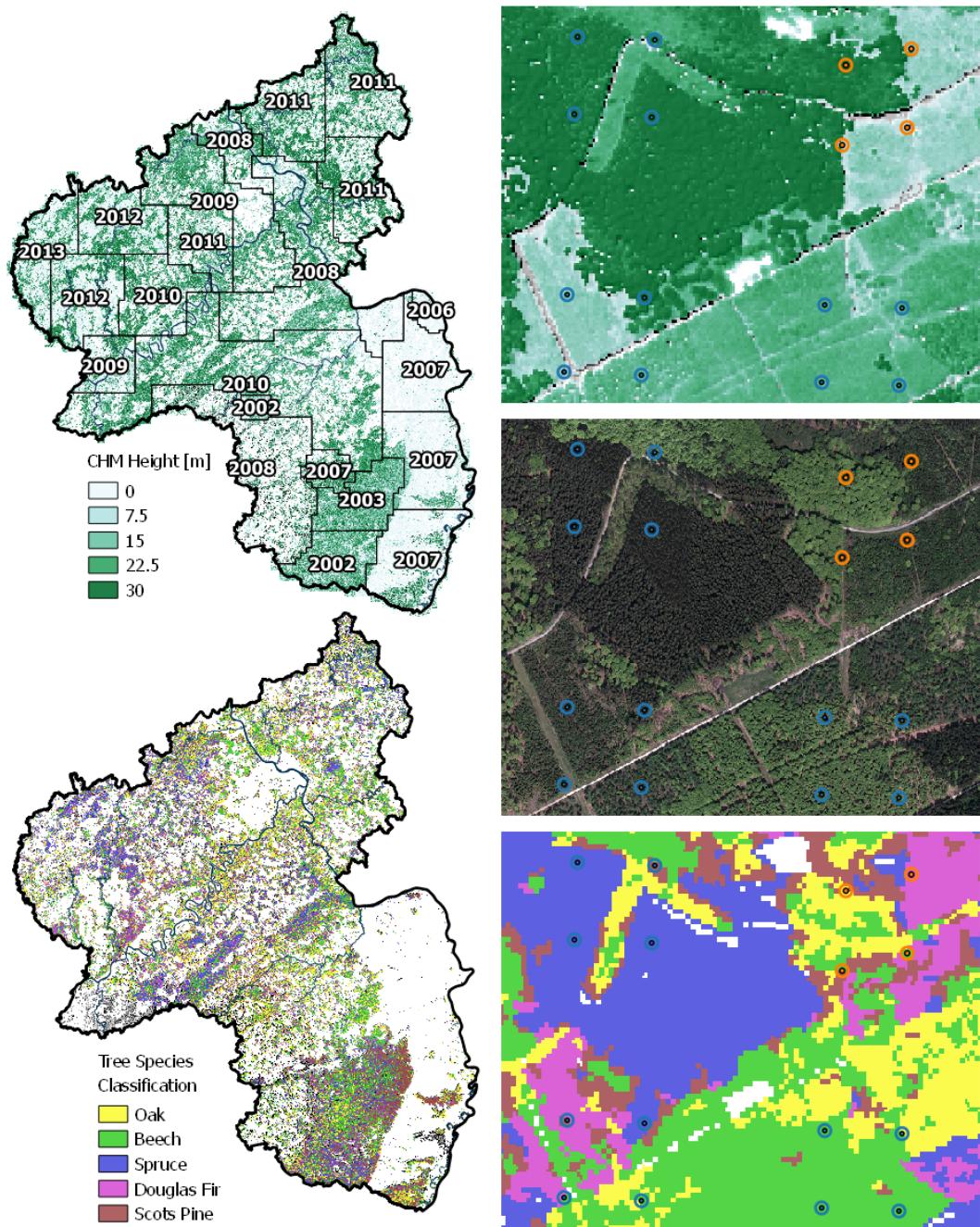


Figure 2: Left: CHM (top) and tree species classification map (bottom) available on the federal state level were used as auxiliary information. Right: Magnified illustration of the supports used to derive the explanatory variables from the auxiliary data.

474 5.5 Calculation of the explanatory variables**475 5.5.1 ALS canopy height model**

476 Continuous explanatory variables derived from the CHM were the mean canopy height
477 (*meanheight*) and the standard deviation (*stddev*). The quantities were calculated by evaluat-
478 ing the raster values around each sample location within a predefined circle (*support*) with a
479 radius of 12 meters. In order to correct for edge effects at the forest border, each support area
480 was previously intersected with the state and communal forest area, which was defined by a
481 polygon mask provided by the state forest service. The percentage of the support within the
482 forest layer was used as the weight $w(x_l)$ introduced in Eq. 10 in order to derive the weighted
483 mean of the explanatory variables on the cluster level. Restricting the auxiliary data evaluation
484 to the forest area was primarily a means to optimize the coherence between explanatory vari-
485 ables computed at the forest boundary and the corresponding local density, but also supports
486 the consistency with the sampling frame (section 5.3). In particular, the BWI3 survey applies
487 an edge correction at the forest border at the individual tree level by intersection of the trees'
488 inclusion circle with the forest border. An adjustment for the fact that part of a trees' inclusion
489 circle is outside the forest area is then realized by increasing its counting factor, which would
490 otherwise lead to an underestimation of the local density. Since the ALS height values will
491 usually drop to around zero outside the forest area, neglecting the edge correction of the sup-
492 port attenuates the mean canopy height value towards zero and thus increases the discrepancy
493 to the local density.

494 We further derived the year of the ALS acquisition (*ALSpyear*) as an additional categorical
495 variable which was used to account for the time-lag between the CHM information and the
496 terrestrial survey as well as to explain heterogeneity in the data introduced by the varying ALS
497 quality. Adjustments were made to the original acquisition years by introducing an additional
498 factor level *2008_I* for a subset of the 2008 acquisition where the quality turned out to be
499 considerably poor due to a sensor error. In addition, the years 2006 and 2007 as well as
500 2012 and 2013 were pooled in order to increase the number of observations per factor level,
501 resulting in nine categories in total (2002, 2003, 2007, 2008_I, 2008, 2009, 2010, 2011 and
502 2012).

503 5.5.2 Tree species map

504 The tree species map information was used to predict the main tree species of the sample
505 trees at each sample plot (*treespecies*) as an additional categorical variable. This implied two
506 consecutive processing steps. In the *first* step, one of the five tree species was assigned to a
507 sample location if 100% of the raster values within the edge-corrected support were classified
508 as that species. Otherwise, the sample location was assigned the value 'mixed'. Likewise for
509 the CHM variables, the support radius was 12 meters although the use of different support
510 sizes for each explanatory variable would be in agreement with the two-phase estimators pre-
511 sented in section 4.2. When using the *treespecies* variable in a regression model, the support
512 size and the percentage threshold particularly constitute parameters to be optimized in order

513 to achieve an optimal variance decomposition of the data that subsequently leads to the best
 514 possible model accuracy. A detailed analysis and description of the optimal processing parameters
 515 for the explanatory variables of the present data set is provided in [Hill et al \(2018\)](#). In
 516 this study, the authors also applied a calibration model to the initially derived *treespecies* vari-
 517 able that successfully removed the effects of misclassification errors on the regression model
 518 coefficients and increased the model accuracy. The calibration model consists of a decision
 519 tree from a random forest algorithm [Breiman \(2001\)](#) that was trained to predict the actual
 520 main plot tree species (known for all terrestrial plots) based on available auxiliary variables.
 521 These variables were the predicted *treespecies* variable, the mean canopy height and standard
 522 deviation of the CHM, as well as the proportion of coniferous trees estimated from the classi-
 523 fication map and the growing region derived from a polygon map. The algorithm was grown
 524 with 2000 trees considering $\sqrt{p} \approx 3$ of the predictors for each split. In a *second* step, we thus
 525 applied this calibration model to the *treespecies* variable derived at all sample locations s_1 .
 526 Table 3 gives the classification accuracies of the *treespecies* variable after calibration.

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

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

527 5.6 Regression Model

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

535 The model with highest adjusted R^2 and lowest RMSE was achieved with the variables
 536 *meanheight*, *meanheight*², *stddev* and *ALSyear* derived from the ALS data as well as the
 537 *treespecies* variable derived from the classification map as main effect terms. Additionally,
 538 interaction terms between *meanheight* and *ALSyear*, *stddev* and *ALSyear*, *meanheight* and
 539 *stddev*, and *meanheight* and *treespecies* were included. The model yielded an adjusted R^2 of
 540 0.48 and an RMSE of 140.62 m³/ha (46.7%) on the plot level (table 4, full model). While the

5. CASE STUDY

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 (formulas given in Appendix) and found the model accuracies to be higher than on the plot level (adjusted R^2 of 0.59 and RMSE of $101.61 \text{ m}^3/\text{ha}$ 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 a highly unbalanced dataset when stratifying according to the *treespecies* variable. For this reason, a species individual modeling within each *ALSyear* stratum remained infeasible, but might have further improved the model accuracy. An additional evaluation of the model residuals within each ALS acquisition year stratum revealed that the model accuracies substantially varied between the strata (table 5). Values above the overall adjusted R^2 were particularly achieved in ALS acquisition years close to the terrestrial survey date (0.60 to 0.65 on the cluster level). This effect was substantially diminished when dropping the *ALSyear* 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 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 +	full model	0.58	90.11	29.76
treespecies + ALSyear +		(0.48)	(139.22)	(45.98)
meanheight:treespecies +				
meanheight:ALSyear + meanheight:stddev +				
stddev:ALSyear				
meanheight + stddev + meanheight ² +	reduced model	0.55	95.23	31.65
ALSyear + meanheight:ALSyear +		(0.45)	(144.13)	(47.60)
meanheight:stddev + stddev:ALSyear				

Concerning the treatment of outliers or leverage points for model fitting, it should be noted that there is a major difference between the external and internal model approach. In order to ensure the zero-mean-residual property of the PSMALL and EXTPSYNT estimator under the internal model approach (section 4.2.1 and 4.2.3), all observations of the terrestrial sample s_2 used for estimation must also be included in the modeling frame. Thus, excluding an

5. CASE STUDY

observation from the model fit does also require its deletion 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 measurement errors). 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. With respect to these considerations, we conducted an analysis of influential observations (Fahrmeir et al, 2013, pp. 160–167) on the plot level for the full regression model. In order to identify implausible observations in the explanatory data, we calculated the *leverage* values and found that the critical threshold of $2p/n$ (i.e. twice the average of the hat matrix' diagonal entries) was exceeded by 10% of all observations. Further investigation revealed that several leverage points showed unusually large *meanheight* values associated to their respective timber volume density values. These implausible height values particularly occurred in ALSyear acquisition years differing from the terrestrial survey date and were thus much more likely caused by harvesting activities in the sample plot area than erroneous ALS data. A deletion of these plots from the sampling frame was thus not justified due to the reasons given before. Also the remaining leverage points could not be definitely attributed to erroneous auxiliary data and were kept in the data set. The same was true for 10 observation with exceptionally large local density values that were identified as *outliers*.

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

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

588 6 Results

589 6.1 General estimation results

590 An application of the SRS, PSMALL and EXTPSYNTH estimator was not feasible for
 591 17 of all 405 FR-units due to an insufficient terrestrial sample size of $n_{2,G} < 2$. We further
 592 restricted the calculation of the PSMALL and EXTPSYNTH estimator to small area units with
 593 a minimum terrestrial sample size of $n_{2,G} \geq 4$ to avoid unstable estimates. This affected 65
 594 additional FR units and limited unbiased two-phase estimations to 321 (79%) of the 405 FR
 595 units. It should be noted that also the PSYNTH estimator could not be applied for 2 FR-units
 596 since $n_{1,G} < 2$. Due to substantial larger sample sizes, all estimators could however be applied
 597 to all 45 FA units. The average value and the range of the mean timber volume estimates
 598 over the evaluated FA and FR units turned out to be very similar between all estimators (table
 599 6). An additional pairwise comparison of the 95% confidence intervals revealed that the four
 600 estimators did in fact not produce statistically different point estimates for all FA and FR units.
 601 This confirmed that the differences between the estimators are solely found in the precision
 602 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

603 6.2 Estimation errors

604 On both small area levels, the design-unbiased estimators PSMALL and EXTPSYNTH led
 605 to a substantial reduction in the estimation errors compared to the SRS estimator (fig. 3). On
 606 the FA level, the SRS estimator yielded an estimation error of 6.7% on average compared to
 607 5.2% and 4.8% under the EXTPSYNTH and PSMALL estimator (table 6). The cumulative
 608 error distribution (fig. 3, left) reveals that under the SRS estimator, errors $\leq 5\%$ were achieved
 609 for 17% of the FA units (8 of 45). This proportion could be increased to 62% (28 FA units) and

6. RESULTS

610 73% (33 FA units) by application of the PSMALL and EXTPSYNTH estimator. 95% of all
 611 estimates exhibited errors $\leq 9.5\%$ under the SRS estimator and $\leq 6.6\%$ when using PSMALL
 612 or EXTPSYNTH. Estimation errors higher than 10% only appeared twice for each of the three
 613 estimators.

614 The error reduction by the PSMALL and EXTPSYNTH estimator was even more pro-
 615 nounced on the FR level (fig. 3, right), although the overall error niveau was substantially
 616 higher than on FA level. The average error under the SRS estimator was 18.3%, while it was
 617 11.3% and 12.2% under the PSMALL and EXTPSYNTH estimator (table 6). Errors smaller
 618 than 10% were achieved for 15% of the FR units by the SRS estimator, and for 46% by the PS-
 619 MALL and PSYNTH estimator. 95% of the 321 FR units where PSMALL and EXTPSYNTH
 620 could be applied exhibited errors $\leq 20\%$. In comparison, the SRS estimates resulted in errors
 621 $\leq 36.6\%$ for 95% of the 388 FR units.

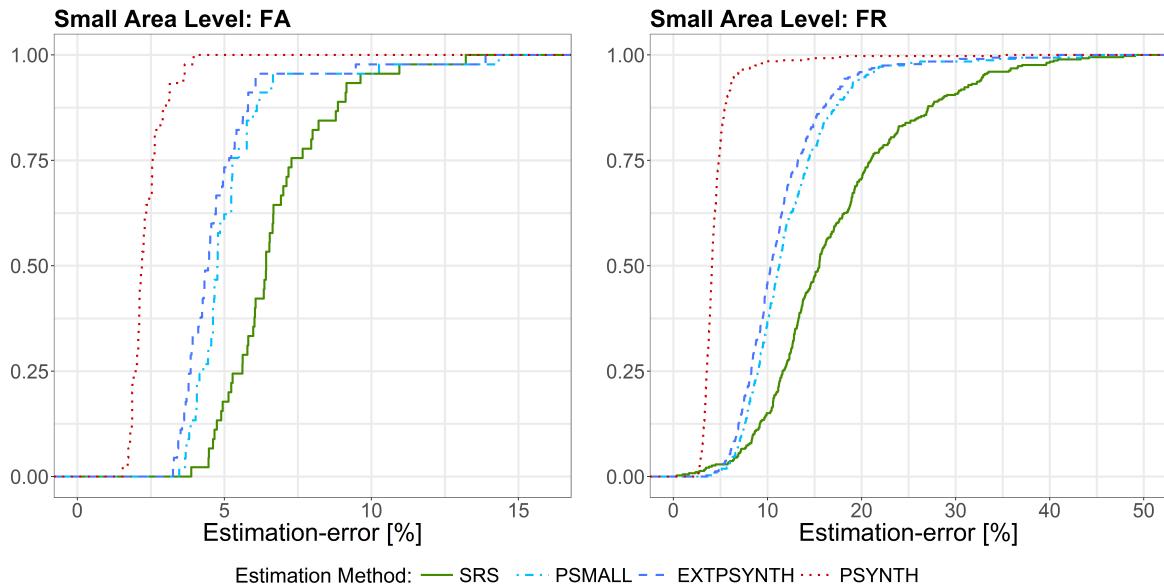


Figure 3: Cumulative distribution of estimation errors under the simple random sampling (SRS), the pseudo small (PSMALL), the extended pseudo synthetic (EXTPSYNTH) and the pseudo synthetic (PSYNTH) estimator. *Left:* Results for the 45 FA units. *Right:* Results for the 388 (SRS), 321 (PSMALL / EXTPSYNTH)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

622 On both small area levels, the PSYNTH estimator resulted in much smaller estimation errors
 623 compared to PSMALL and EXTPSYNTH. This was as expected, since the PSYNTH variance
 624 estimate does not take the residual variation in each small area unit into account (section 4.2.2).
 625 Compared to the asymptotically design-unbiased estimators PSMALL and EXTPSYNTH, the
 626 estimation errors produced by PSYNTH thus seem to be too optimistic. One should also recall
 627 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.

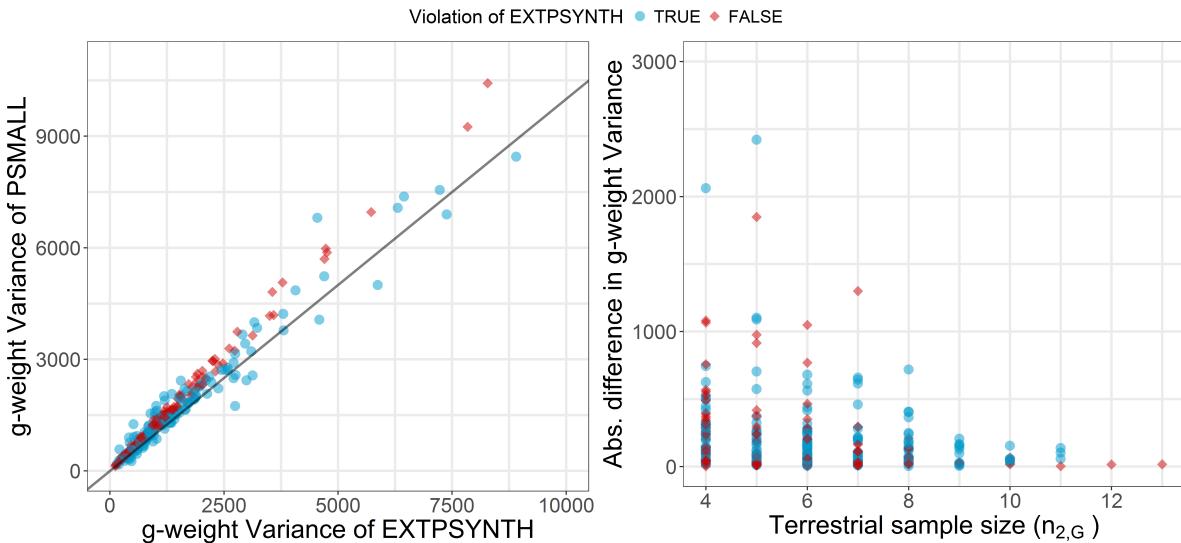


Figure 4: *Left:* Comparison of the g-weight variance between the PSMALL and the EXTPSYNTH estimator for the 321 FR units. *Right:* Difference in g-weight variance between the PSMALL and the EXTPSYNTH estimator in dependence of the terrestrial data (n_{2G}) in the FR unit.

6.4 Reduction of SRS variance by PSMALL and EXTPSYNTH

A direct comparison of the realized variances within the small area units revealed that the application of the design-unbiased estimators (PSMALL and EXTPSYNTH) led to a reduction of the respective 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 and fig. 5). The PSMALL estimator again yielded slightly lower variance reductions and relative efficiencies due to the generally smaller variances of the EXTPSYNTH estimator (section 6.3).

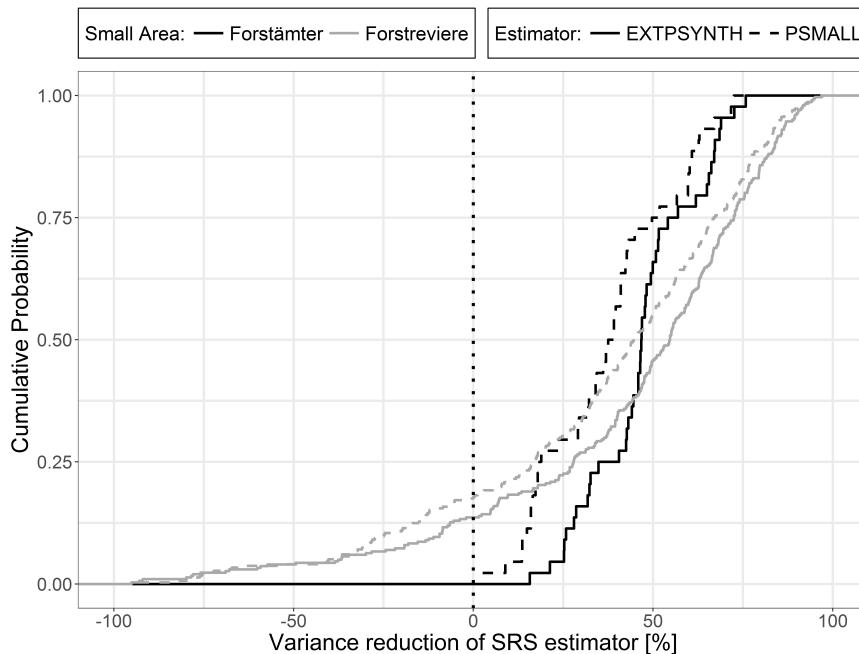


Figure 5: Cumulative distribution of variance reduction by the PSMALL and EXTPSYNTH compared to the SRS estimator for the 45 FA and 321 FR units.

Table 7: Descriptive summary of 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

658 However, cases also occurred on the FR level where one or both two-phase estimators pro-
 659 duced larger variance values than under the SRS estimator. This particularly happened in 19%
 660 (61) of the FR units under the EXTPSYNTH, and in 24% (76) of the FR units under the PS-
 661 MALL estimator. One possible reason for this was supposed to be a large residual variance
 662 due to a poor performance of the regression model within the small area unit. In order to
 663 investigate this hypothesis, we analyzed the three variance terms of the PSMALL estimator
 664 (eq. 14b), i.e. the variance introduced by the uncertainty of the regression coefficients (term
 665 1), the variance caused by estimating the auxiliary means (term 2), and the variance of the
 666 model residual (term 3). The latter particularly expresses the model performance within the
 667 small area unit. Figure 6 illustrates the percentage reduction or increase of the SRS variance
 668 when compared to the PSMALL variance for all FR units in dependence on a) the impact of
 669 residual variance on the g-weight variance, and b) the terrestrial sample size $n_{2,G}$.

670 Obviously, the residual term generally constitutes the dominating part of the PSMALL g-
 671 weight variance (around 84% on average). However, a high proportion of the residual variance
 672 term seems not to be the driver for large PSMALL variances, as apparent from Fig. 6 (right).
 673 The FR units where the PSMALL estimator produced larger variances than the SRS estimator
 674 did not systematically differ from the majority of the FR units where PSMALL performed
 675 better than SRS. However, FR units with exceptionally large variance increases compared to
 676 SRS particularly occurred under small terrestrial sample sizes of $n_{2,G} = 4$ (fig. 6, left). These
 677 FR units exhibited a 272% average increase of the SRS variance, compared to 62% for the
 678 critical FR units with $n_{2,G} > 4$. In comparison, the achievable reduction in SRS variances
 679 compared to SRS were not considerably impacted by the terrestrial sample size (fig. 6, right).
 680 The average reduction of these units was around 50% under sampling sizes of both $n_{2,G} = 4$
 681 and $n_{2,G} > 4$.

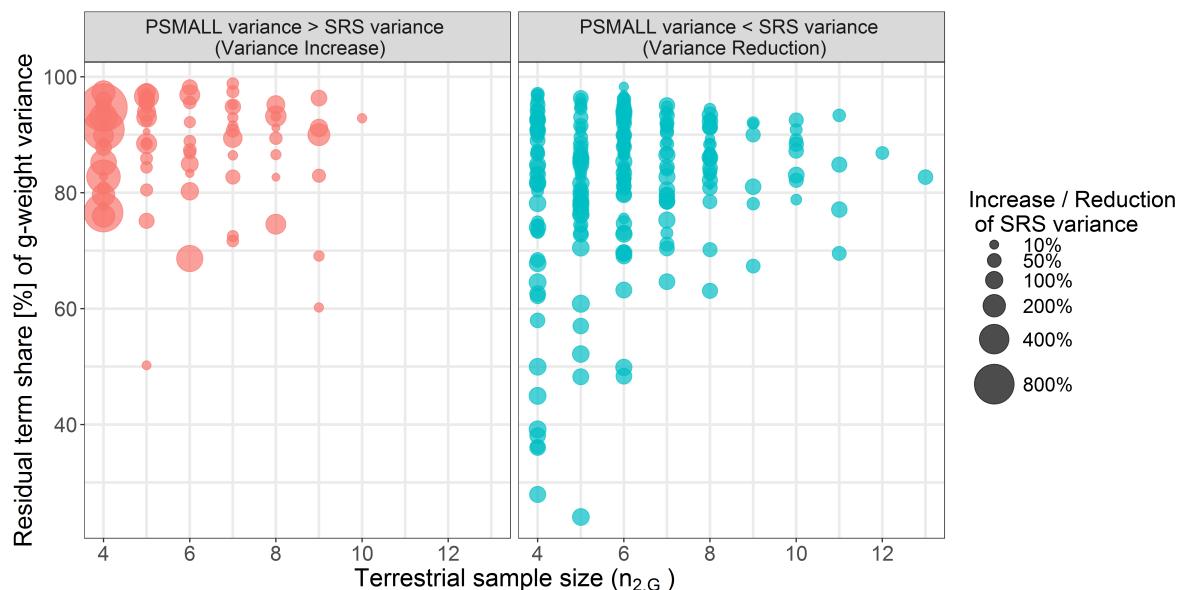


Figure 6: Increase or reduction of the SRS variance by application of the PSMALL estimator for all 321 FR units in dependence of a) the terrestrial sample size in the FR units (x-axis) and b) the residual term proportion of the PSMALL g-weight variance (y-axis).

682 7 Discussion**683 7.1 Performance of estimators**

684 The aim of this study was to investigate the performance of design-based estimators for
685 small area estimation of mean standing timber volume on two spatial forest management levels
686 in Germany. It was of particular interest to gather information about the estimation error levels
687 that can be realized using the German NFI data, characterized by low sampling intensities in
688 the area of interests, as terrestrial phase information in the two-phase estimators. To address
689 these research questions, we applied the simple random sampling (SRS), the pseudo-small
690 (PSMALL) and the extended pseudo-synthetic (EXTPSYNTH) estimator for cluster sampling
691 to 45 and 405 small area units in the German state of Rhineland-Palatinate.

692 Our study showed that on both small area levels, the PSMALL and the EXTPSYNTH esti-
693 mator generally leads to a substantial reduction of the estimation errors produced by the SRS
694 estimator. On the upper management level (FA districts), PSMALL and EXTPSYNTH al-
695 lowed to provide estimation errors smaller than 5% for 73% of the small areas compared to
696 only 17% under the SRS estimator. This level of estimation accuracy could not be achieved on
697 the subsequent management level (FR districts). The reason for this were primarily the sub-
698 stantially smaller terrestrial sample sizes. However, in most FR units (95%), the estimation
699 errors could be limited to 20% compared to 40% under SRS. A pairwise comparison of the
700 confidence intervals revealed that the estimators did thereby not produce significantly differ-
701 ent point estimates. The much smaller estimation errors of the PSYNTH estimator suggested
702 that the neglection of the regression model accuracies in the small area units leads to overop-
703 timistic estimation accuracies and confidence intervals. If possible, one should thus prefer the
704 unbiased estimates of PSMALL or EXTPSYNTH.

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

719 We were also able to empirically confirm that the EXTPSYNTH estimator generally pro-
720 duces slightly smaller variances and estimation errors than the PSMALL. This is most proba-
721 bly caused by marginally smaller model residuals due to the intercept adjustment to the terres-

722 trial data in the small area unit, which is primarily a means to ensure the zero-mean-residual
723 property of the EXTPSYNTH. However, our analysis indicated that the difference between
724 the two estimators are negligible for sample sizes ≥ 10 due to their asymptotic equivalency.
725 Furthermore, one or more clusters not entirely included in the small area unit did reassuringly
726 not have a notable impact on the estimates of EXTPSYNTH. This empirical evidence can
727 however not be generalized to any application of EXTPSYNTH under cluster sampling. It
728 thus seems recommendable to calculate both PSMALL and EXTPSYNTH, and subsequently
729 compare their results. If no suspicious deviations occur, we consider the EXTPSYNTH as the
730 estimator of choice.

731 7.2 Auxiliary data

732 The auxiliary data used in our study were derived from two remote sensing sources, i.e. an
733 ALS canopy height model and a tree species classification map. Likewise in many similar
734 studies, the ALS mean canopy height proved to be the explanatory variable with highest pre-
735 dictive power. However, the large time-gaps of up to 10 years between the ALS acquisition
736 and the terrestrial survey date caused a substantial introduction of artificial noise in the data.
737 Whereas a stratification to the ALS acquisition years was an effective means to counteract the
738 implied residual inflation, several leverage points were unambiguously caused by the temporal
739 asynchronicity. As opposed to the ALS data, the availability of a country-wide tree species
740 classification map has yet been unique among all German federal states. Whereas the study
741 of [Hill et al \(2018\)](#) already showed that the tree species information was able to improve the
742 model accuracy, it has yet not been used to its full potential. One reason for this was the im-
743 possibility of a tree species individual modeling within each ALS acquisition year, which we
744 consider to add further explanatory power. Another reason was the lack of available satellite
745 data for classification in some parts of the country, which led to missingness in the inventory
746 data and restricted 19 FR units to a simpler regression model. Promising steps with respect
747 to more up-to-date canopy height information have already been made, as the topographic
748 survey institution of RLP will from this year on provide a country-wide canopy height model
749 derived from aerial imagery acquisitions. These campaigns will in the future be conducted in
750 a two-year period and allow to derive canopy height information matching the dates of ter-
751 restrial forest inventories. A study of [Kirchhoefer et al \(2017\)](#) recently indicated that similar
752 model accuracies for German NFI data can be achieved using such imagery-based canopy
753 height models. Due to the improved coverage and repetition rate of the Sentinel-2 satellite,
754 also the tree species classification map will in the future be updated each year. We consider
755 these alternative auxiliary data sources to also solve the problem of missing explanatory vari-
756 ables at inventory plots. Thus, one could also make use of the exhaustive information within
757 the two-phase estimators by using the true the auxiliary means ([Mandallaz, 2013a; Mandallaz](#)
758 [et al, 2013](#)), which theoretically leads to a further decrease of the estimation errors. Previous
759 studies of [Mandallaz et al \(2013\)](#) however showed that given a reasonable large sample size
760 of the first phase, the differences in the estimation error are usually small. With respect to the
761 substantial improvements in the temporal synchronicity between auxiliary and terrestrial in-

762 inventory data, we consider the demonstrated double-sampling approach also to be very efficient
763 for change estimation ([Massey and Mandallaz, 2015](#)).

764 **8 Conclusion**

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

781 **Acknowledgements**

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

A Appendix

R-squared on cluster level

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

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

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

RMSE on cluster level

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

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

The *relative* or *normalized* RMSE is calculated by setting the absolute RMSE in relation to the estimated sample mean $\hat{\bar{Y}}_c$:

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

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

803 References

- 804 Bitterlich W (1984) The relascope idea. Relative measurements in forestry. Commonwealth
805 Agricultural Bureaux
- 806 Böckmann T, Saborowski J, Dahm S, Nagel J, Spellmann H (1998) A new conception for
807 forest inventory in lower saxony. Forst und Holz (Germany)
- 808 Breidenbach J, Astrup R (2012) Small area estimation of forest attributes in the norwegian
809 national forest inventory. European Journal of Forest Research 131(4):1255–1267, DOI
810 10.1007/s10342-012-0596-7
- 811 Breiman L (2001) Random forests. Machine learning 45(1):5–32, DOI 10.1023/A:
812 1010933404324
- 813 Bundesministerium für Ernährung LuV (2011) Aufnahmeanweisung für die dritte Bun-
814 deswaldinventur BWI3 (2011 - 2012). URL [https://www.bundeswaldinventur.de/
815 index.php?id=421](https://www.bundeswaldinventur.de/index.php?id=421)
- 816 Fahrmeir L, Kneib T, Lang S, Marx B (2013) Regression: models, methods and applica-
817 tions. Springer Science & Business Media, DOI 10.1007/978-3-642-34333-9, URL <http://www.springer.com/us/book/9783642343322>
- 819 Gauer J, Aldinger E (2005) Waldökologische naturräume deutschlands-wuchsgebiete. Mit-
820 teilungen des Vereins für Forstliche Standortskunde und Forstpflanzenzüchtung 43:281–288
- 821 Goerndt ME, Monleon VJ, Temesgen H (2011) A comparison of small-area estimation tech-
822 niques to estimate selected stand attributes using lidar-derived auxiliary variables. Canadian
823 journal of forest research 41(6):1189–1201
- 824 Grafström A, Schnell S, Saarela S, Hubbell S, Condit R (2017) The continuous population
825 approach to forest inventories and use of information in the design. Environmetrics
- 826 Gregoire TG, Valentine HT (2007) Sampling strategies for natural resources and the environ-
827 ment. CRC Press
- 828 Hill A, Massey A (2017) forestinventory: Design-Based Global and Small-Area Estima-
829 tions for Multiphase Forest Inventories. R package version 0.3.1. URL [https://cran.
830 r-project.org/web/packages/forestinventory/index.html](https://cran.r-project.org/web/packages/forestinventory/index.html)
- 831 Hill A, Buddenbaum H, Mandallaz D (2018) Combining canopy height and tree species map
832 information for large scale timber volume estimations under strong heterogeneity of auxil-
833 iary data and variable sample plot sizes, submitted for publication
- 834 Kirchhoefer M, Schumacher J, Adler P, Kändler G (2017) Considerations towards a novel ap-
835 proach for integrating angle-count sampling data in remote sensing based forest inventories.
836 Forests 8(7):239

- 837 Koch B (2010) Status and future of laser scanning, synthetic aperture radar and hyperspectral
838 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
839 <https://doi.org/10.1016/j.isprsjprs.2010.09.001>
- 841 Köhl M, Magnussen SS, Marchetti M (2006) Sampling methods, remote sensing and GIS
842 multiresource forest inventory. Springer Science & Business Media
- 843 Kublin E, Breidenbach J, Kändler G (2013) A flexible stem taper and volume prediction
844 method based on mixed-effects b-spline regression. European journal of forest research
845 132(5-6):983–997, DOI 10.1007/s10342-013-0715-0
- 846 Kuliešis A, Tomter SM, Vidal C, Lanz A (2016) Estimates of stem wood increments in forest
847 resources: comparison of different approaches in forest inventory: consequences for inter-
848 national reporting: case studyof european forests. Annals of Forest Science 73(4):857–869
- 849 Lamprecht S, Hill A, Stoffels J, Udelhoven T (2017) A machine learning method for co-
850 registration and individual tree matching of forest inventory and airborne laser scan-
851 ning data. Remote Sensing 9(5), DOI 10.3390/rs9050505, URL <http://www.mdpi.com/2072-4292/9/5/505>
- 853 von Lüpke N (2013) Approaches for the optimisation of double sampling for stratification in
854 repeated forest inventories. PhD thesis, University of Göttingen
- 855 von Lüpke N, Hansen J, Saborowski J (2012) A three-phase sampling procedure for contin-
856 uous forest inventory with partial re-measurement and updating of terrestrial sample plots.
857 European Journal of Forest Research 131(6):1979–1990, DOI 10.1007/s10342-012-0648-z
- 858 LWaldG (2000) Landeswaldgesetz Rheinland-Pfalz (Forest Act Rhineland-Palatinate). URL
859 <http://www.landesrecht.rlp.de>, Rhineland-Palatinate, Germany
- 860 Magnussen S, Mandallaz D, Breidenbach J, Lanz A, Ginzler C (2014) National forest inven-
861 tories in the service of small area estimation of stem volume. Canadian Journal of Forest
862 Research 44(9):1079–1090
- 863 Magnussen S, Mauro F, Breidenbach J, Lanz A, Kändler G (2017) Area-level analysis of forest
864 inventory variables. European Journal of Forest Research pp 1–17
- 865 Mandallaz D (2008) Sampling techniques for forest inventories. CRC Press, DOI 10.1201/
866 9781584889779, URL <https://doi.org/10.1201/9781584889779>
- 867 Mandallaz D (2013a) Design-based properties of some small-area estimators in forest inven-
868 tory with two-phase sampling. Canadian Journal of Forest Research 43(5):441–449, DOI
869 10.1139/cjfr-2012-0381, URL <https://doi.org/10.1139/cjfr-2012-0381>
- 870 Mandallaz D (2013b) A three-phase sampling extension of the generalized regression estima-
871 tor with partially exhaustive information. Canadian Journal of Forest Research 44(4):383–
872 388, DOI 10.1139/cjfr-2013-0449, URL <https://doi.org/10.1139/cjfr-2013-0449>

- 873 Mandallaz D, Breschan J, Hill A (2013) New regression estimators in forest inventories
874 with two-phase sampling and partially exhaustive information: a design-based monte
875 carlo approach with applications to small-area estimation. Canadian Journal of Forest
876 Research 43(11):1023–1031, DOI 10.1139/cjfr-2013-0181, URL <https://doi.org/10.1139/cjfr-2013-0181>
- 878 Mandallaz D, Hill A, Massey A (2016) Design-based properties of some small-area estima-
879 tors in forest inventory with two-phase sampling - revised version. Tech. rep., Department
880 of Environmental Systems Science, ETH Zurich, DOI 10.3929/ethz-a-010579388, URL
881 <https://doi.org/10.3929/ethz-a-010579388>
- 882 Massey A, Mandallaz D (2015) Design-based regression estimation of net change for for-
883 est inventories. Canadian Journal of Forest Research 45(12):1775–1784, DOI 10.1139/
884 cjfr-2015-0266, URL <https://doi.org/10.1139/cjfr-2015-0266>, <https://doi.org/10.1139/cjfr-2015-0266>
- 886 Massey A, Mandallaz D, Lanz A (2014) Integrating remote sensing and past inventory data
887 under the new annual design of the swiss national forest inventory using three-phase design-
888 based regression estimation. Canadian Journal of Forest Research 44(10):1177–1186, DOI
889 10.1139/cjfr-2014-0152, URL <https://doi.org/10.1139/cjfr-2014-0152>
- 890 Naesset E (2014) Area-based inventory in norway – from innovation to an operational reality.
891 In: Forest Applications of Airborne Laser Scanning - Concepts and Case Studies, Springer,
892 chap 11, pp 216–240, DOI 10.1007/978-94-017-8663-8
- 893 Polley H, Schmitz F, Hennig P, Kroher F (2010) Germany. In: National Forest Invento-
894 ries - Pathways for Common Reporting, Springer, chap 13, pp 223–243, DOI 10.1007/
895 978-90-481-3233-1
- 896 Rao JN (2015) Small-Area Estimation. Wiley Online Library
- 897 Saborowski J, Marx A, Nagel J, Böckmann T (2010) Double sampling for stratification
898 in periodic inventories—finite population approach. Forest ecology and management
899 260(10):1886–1895
- 900 Särndal CE, Swensson B, Wretman J (2003) Model assisted survey sampling. Springer Sci-
901 ence & Business Media
- 902 Schmitz F, Polley H, Hennig P, Dunger K, Schwitzgebel F (2008) Die zweite Bundeswaldin-
903 ventur - BWI2: Inventur- und Auswertmethoden, Bundesministerium fur Ernährung, Land-
904 Wirtschaft und Verbraucherschutz (Hrsg)
- 905 Steinmann K, Mandallaz D, Ginzler C, Lanz A (2013) Small area estimations of proportion
906 of forest and timber volume combining lidar data and stereo aerial images with terrestrial
907 data. Scandinavian journal of forest research 28(4):373–385

- 908 Stoffels J, Mader S, Hill J, Werner W, Ontrup G (2012) Satellite-based stand-wise forest cover
909 type mapping using a spatially adaptive classification approach. European journal of forest
910 research 131(4):1071–1089
- 911 Stoffels J, Hill J, Sachtleber T, Mader S, Buddenbaum H, Stern O, Langshausen J, Dietz
912 J, Ontrup G (2015) Satellite-based derivation of high-resolution forest information layers
913 for operational forest management. Forests 6(6):1982–2013, DOI 10.3390/f6061982, URL
914 <http://www.mdpi.com/1999-4907/6/6/1982>
- 915 Thünen-Institut (2014) Dritte Bundeswaldinventur 2012. URL <https://bwi.info>, accessed:
916 2017-02-03
- 917 White JC, Coops NC, Wulder MA, Vastaranta M, Hilker T, Tompalski P (2016) Remote
918 sensing technologies for enhancing forest inventories: A review. Canadian Journal of Re-
919 mote Sensing 42(5):619–641, DOI 10.1080/07038992.2016.1207484, URL <http://dx.doi.org/10.1080/07038992.2016.1207484>