

# Article Draft

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

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

Andreas Hill<sup>\*†</sup>, Daniel Mandallaz<sup>\*</sup>, Joachim Langshausen<sup>\*\*</sup>

<sup>\*</sup>Department of Environmental Systems Science, ETH Zurich

\*\*Forest Service Rhineland-Palatinate

<sup>10</sup> Corresponding author<sup>†</sup>: [andreas.hill@usys.ethz.ch](mailto:andreas.hill@usys.ethz.ch)  
<sup>11</sup> Co-authors: [daniel.mandallaz@env.ethz.ch](mailto:daniel.mandallaz@env.ethz.ch), [joachim.langshausen@wald-rlp.de](mailto:joachim.langshausen@wald-rlp.de)

<sup>12</sup> Publication date: March 1, 2018

<sup>13</sup> **Abstract**

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

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

# 33 Contents

34	<b>1 Introduction</b>	1
35	<b>2 Terrestrial sampling design of the German NFI</b>	3
36	<b>3 Double sampling in the infinite population approach</b>	3
37	<b>4 Estimators</b>	4
38	4.1 Design-based one-phase estimator for cluster sampling (SRS) . . . . .	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	<b>5 Case study</b>	11
45	5.1 Study area and small area units . . . . .	11
46	5.2 Terrestrial sample . . . . .	11
47	5.3 Extension to double-sampling design . . . . .	12
48	5.4 Auxiliary data . . . . .	13
49	5.4.1 LiDAR canopy height model . . . . .	13
50	5.4.2 Tree species map . . . . .	14
51	5.5 Calculation of the explanatory variables . . . . .	16
52	5.5.1 Canopy height model . . . . .	16
53	5.5.2 Tree species map . . . . .	16
54	5.6 Regression Model . . . . .	17
55	<b>6 Results</b>	20
56	6.1 General estimation results . . . . .	20
57	6.2 Estimation error . . . . .	20
58	6.3 Comparison of PSMALL and EXTPSYNTH . . . . .	22
59	6.4 Variance reduction compared to SRS . . . . .	23
60	<b>7 Discussion</b>	25
61	7.1 Performance of estimators . . . . .	25
62	7.2 Auxiliary data . . . . .	26
63	<b>8 Conclusion</b>	27
64	<b>A Appendix</b>	28

65 **List of Figures**

66 1	Left: Study area with delineated FA forest management units. Right: Example 67 for each of the three management units (from top to bottom): FA, FR and 68 forest stand unit overlayed with the extended double-sampling cluster design. 69 <i>Green:</i> Forest stand polygon layer defining the forest area of this study. . . . .	12
70 2	Left: CHM (top) and tree species classification map (bottom) available on the 71 federal state level. Right: Magnified illustration of the supports used to derive 72 the explanatory variables from the auxiliary data. . . . .	15
73 3	Cumulative distribution of estimation errors under SRS, PSMALL, EXTP- 74 SYNTH and the PSYNTH estimator. <i>Left:</i> Results for the 45 FA units. 75 <i>Right:</i> Results for the 388 (SRS), 321 (PSMALL, EXTPSYNTH) and 403 76 (PSYNTH) FR units. . . . .	21
77 4	<i>Left:</i> Comparison of the g-weight variance between the PSMALL and the 78 EXTPSYNTH estimator for the 321 FR units. <i>Right:</i> Difference in g-weight 79 variance between the PSMALL and the EXTPSYNTH estimator in depen- 80 dence of the terrestrial data ( $n_{2,G}$ ) in the FR unit. . . . .	22
81 5	Cumulative distribution of variance reduction by the PSMALL and EXTP- 82 SYNTH compared to the SRS estimator for the 45 FA and 321 FR units. . . .	23
83 6	Share of the overall variance by the residual term of the PSMALL estimator 84 for various small area sample sizes. Points are scaled by the overall percentage reduction/increase of the variance compared to SRS. . . . .	24

86 **List of Tables**

87 1	Descriptive statistics of the forest observed on NFI sample plots located within communal and state forest area (n=5791). . . . .	13
88 2	Sample size for each phase in entire study area. $n_{\{1,2\},plots}$ : number of plots. $n_{\{1,2\}}$ : number of clusters. TSPEC: tree species map information. . . . .	14
89 3	Classification accuracies of the <i>treespecies</i> variable before and after calibra- tion. $n_{ref}$ : number of terrestrial reference plots. $n_{class}$ : number of classified plots. . . . .	17
90 4	Model fit metrics for the two OLS regression models on the cluster level. In- teraction terms are indicated by ':'. () give the respective values on the plot level. . . . .	18
91 5	$R^2$ , RMSE and RMSE% on the cluster level of the full regression model within ALS acquisition year strata ( <i>ALSpyear</i> ). <i>Area<sub>ALSpyear</sub></i> : Area covered by ALS ac- quisition given in km <sup>2</sup> . <i>n</i> : sample size of validation data. () give the respective values on the plot level. . . . .	19
92 6	Descriptive summary of point estimates and estimation errors on the two forest district levels. $N_u$ : number of evaluated small area units. . . . .	20
93 7	Descriptive summary of variance reduction compared to SRS and relative ef- ficiencies on the two forest district levels. $N_u$ : number of evaluated small area units. . . . .	23
94		
95		
96		
97		
98		
99		
100		
101		
102		
103		
104		
105		

<sup>106</sup> **1 Introduction**

<sup>107</sup> The German National Forest Inventory (NFI) provides reliable evidence-based and accu-  
<sup>108</sup> rate information of the current state and the development of Germany's forest over time. The  
<sup>109</sup> NFI thereby has the responsibility to satisfy various information needs including reporting to  
<sup>110</sup> public and state forestry administrations, wood-based industries and the public on the national  
<sup>111</sup> level, as well as to the Food and Agriculture Organization of the United Nations (FAO) and to  
<sup>112</sup> the United Nations Framework Convention on Climate Change (UNFCCC) on the international  
<sup>113</sup> level ([Polley et al, 2010](#)). The current design of the German NFI rests solely upon a terrestrial  
<sup>114</sup> cluster inventory that is carried out at sample locations systematically distributed over the en-  
<sup>115</sup> tire forest state area of Germany. In order to cover a large area of 114'191 ha ([Thünen-Institut,](#)  
<sup>116</sup> [2014](#)), the sample size has been specifically chosen to satisfy high estimation accuracies for  
<sup>117</sup> forest attributes on the national and federal state levels. However, sample sizes often drop  
<sup>118</sup> dramatically when entering spatial units below the federal state level. This is particularly true  
<sup>119</sup> for forest management levels such as forest districts for which the estimation uncertainties  
<sup>120</sup> turn out to be unacceptably large due to the very limited number of sample plots within these  
<sup>121</sup> units. For this reason, the German NFI data have not yet been extensively incorporated into  
<sup>122</sup> operational planning on forest district management levels. In most German federal states,  
<sup>123</sup> management strategies are thus still based on expert judgments from time-consuming stand-  
<sup>124</sup> wise inventories (SFI), which are prone to systematic deviations [Kuliešis et al \(2016\)](#) and do  
<sup>125</sup> not provide any measure of uncertainty.

<sup>126</sup> Some German federal states, such as Lower Saxony, have approached this problem by es-  
<sup>127</sup> tablishing a regional Forest District Inventory (FDI) with a much higher sampling density  
<sup>128</sup> than used by the NFI in order to scientifically base their regional management strategies on  
<sup>129</sup> quantitative and accurate information ([Böckmann et al, 1998](#)). However, such FDIs are cost-  
<sup>130</sup> intensive and, facing increasing restrictions in budget and staff resources, there has been a  
<sup>131</sup> need for more cost-efficient inventory methods ([von Lüpke, 2013](#)). One method which has  
<sup>132</sup> proven to be efficient is double- or two-phase sampling ([Särndal et al, 2003; Gregoire and](#)  
<sup>133</sup> [Valentine, 2007; Köhl et al, 2006; Mandallaz, 2008](#)). Double-sampling incorporates less ex-  
<sup>134</sup> pensive auxiliary information and can be used to either increase estimation precision under a  
<sup>135</sup> fixed terrestrial sample size, or maintain estimation precision under reduced terrestrial sam-  
<sup>136</sup> ple size. Double-sampling procedures have already been used for stratification in the FDI of  
<sup>137</sup> Lower Saxony ([Saborowski et al, 2010](#)), and [Grafström et al \(2017\)](#) illustrated how to use  
<sup>138</sup> the auxiliary information to determine optimised balanced terrestrial sample designs. Recent  
<sup>139</sup> studies have extended double-sampling to triple-sampling estimation methods using auxiliary  
<sup>140</sup> information derived at two different sampling intensities. An example can be found in [von](#)  
<sup>141</sup> [Lüpke et al \(2012\)](#) who illustrated an extension of the existing two-phase FDI of Lower Sax-  
<sup>142</sup> ony to a three-phase design that uses updates of past inventory data as additional auxiliary  
<sup>143</sup> information and allows for a significant reduction of the terrestrial sample size in intermediate  
<sup>144</sup> inventories. Another example is [Massey et al \(2014\)](#) who developed a triple-sampling exten-  
<sup>145</sup> sion based on the ideas of [Mandallaz \(2013b\)](#) for the Swiss NFI that can significantly reduce  
<sup>146</sup> the increase in estimation uncertainty caused by the new annual inventory design.

<sup>147</sup> Two-phase and three-phase samplings techniques have also been applied to small area es-

## 1. INTRODUCTION

---

timation (SAE). SAE techniques address the situation where the number of samples within a subunit, or small area (SA), of the entire sampling frame is too small to provide reliable estimates for that unit. A broad range of SA estimators used in forest inventories (Köhl et al, 2006) originally comes from official statistics. One such method that is commonly applied is 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 observed in the small area. These models are trained using data from outside the small area in order to "borrow strength" from areas where information is available. Of numerous applications of SAE in forestry (Breidenbach and Astrup, 2012; Goerndt et al, 2011; Steinmann et al, 2013; Mandallaz et al, 2013), most use unit-level models, i.e. the inventory plot is the unit of the response variable in the training data used for the model fit. Such unit-level models have been intensively investigated for timber volume estimation using various remote sensing auxiliary data (Koch, 2010; Naesset, 2014). Other studies 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 derived under low sampling densities can still be used to provide acceptable precision of small area estimates on much smaller management levels. One example is Breidenbach and Astrup (2012) who used data from the Norwegian NFI to make small area estimation for standing timber volume for 14 municipalities where the number of NFI samples within these areas were between 1 and 35. The estimation errors under the applied model-dependent and design-based small area estimators turned to be markedly smaller than under the standard one-phase estimator. Another example is Magnussen et al (2014) who recently used the Swiss NFI data to estimate timber volume within 108 Swiss forest districts with sample sizes between 9 and 206. Similar studies using German NFI data for small area estimation have been lacking.

The aim of this study was to investigate whether the German NFI data can provide acceptable estimation precision on two forest district levels using the latest small area estimation procedures. We therefore conducted a study in the German federal state Rhineland-Palatinate where we extended the German NFI to a double-sampling design and applied three types of design-based small area regression estimators in order to derive point and variance estimates of mean standing timber volume for 45 and 405 forest districts respectively. The SA-estimators we considered were the *pseudo-small*, *extended pseudo-synthetic* and the *pseudo-synthetic* design-based small area estimator suggested by Mandallaz (2013a) and Mandallaz et al (2013). Auxiliary data consisted of a canopy height model (CHM) obtained from a countrywide airborne laser scanning (ALS) and a tree species classification map to be used for regression within tree species strata. The estimation precisions were compared to those obtained by the standard one-phase estimator for cluster sampling under simple random sampling. The chosen double-sampling estimators were selected for several reasons: (i) the design-based framework relaxes dependencies on the regression model assumptions which seemed appropriate facing severe quality restrictions in the ALS data; (ii) the estimators can be used with *non-exhaustive*, i.e. non wall-to-wall, auxiliary information; (iii) all estimators are explicitly formulated for cluster sampling which has not yet been the case for frequently used model-dependent estimators; and (iv) the asymptotically unbiased g-weight variance accounts for estimating the regression coefficients on the same sample used for estimation (*internal model approach*) and is also robust under heteroscedasticity of the model residuals. The results from

192 this study were considered to provide valuable information whether the suggested procedure  
193 might be a cost-saving alternative to a regional FDI.

## 194 **2 Terrestrial sampling design of the German NFI**

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

## 214 **3 Double sampling in the infinite population approach**

215 The estimators used in this study have been proposed by ([Mandallaz, 2013a; Mandallaz](#)  
216 [et al, 2013](#)) and derive their mathematical properties under the so-called infinite population  
217 approach. Therefore, we shall first provide a short introduction into this general estimation  
218 framework. We start by assuming that the population  $P$  of trees  $i \in 1, 2, \dots, N$  within a forest of  
219 interest  $F$  is exactly defined, and each tree  $i$  has a response variable  $Y_i$  (e.g. its timber volume)  
220 that can be used to define the population mean  $Y$  (e.g., the average timber volume per unit area)  
221 over  $F$ . Since a full census of all tree population individuals is almost never feasible,  $Y$  has  
222 to be estimated based on a sample. In the infinite population approach this sample is a set of  
223 points or locations  $x$  distributed independently and uniformly over the set of all possible points  
224 in  $F$ . Each point  $x$  has an associated local density  $Y(x)$  (e.g., the timber volume per unit area)  
225 whose spatial distribution is given by a fixed (i.e. non stochastic) piecewise constant func-  
226 tion. The population mean  $Y$  is mathematically equivalent to the integral of the local density  
227 function surface divided by the surface area of  $F$ ,  $\lambda(F)$ , i.e.  $Y = \frac{1}{N} \sum_{i=1}^N Y_i = \frac{1}{\lambda(F)} \int_F Y(x) dx$ ,

and thus the population mean  $Y$  corresponds to a spatial mean. Since the actual local density function is unobserved in its entirety, one estimates  $Y$  by taking a sample  $s_2$  consisting of  $n_2$  points and measuring each of their respective local densities. This sampling procedure is often referred to as *one-phase sampling* (OPS) and  $s_2$  is referred to as the terrestrial inventory. In contrast to the one-phase approach, *two-phase* or *double-sampling* procedures use information from two nested samples (phases). Practically speaking, the terrestrial inventory  $s_2$  is embedded in a large phase  $s_1$  comprising  $n_1$  sample locations that each provide a set of explanatory variables described by the column vector  $\mathbf{Z}(x) = (z(x)_1, z(x)_2, \dots, z(x)_p)^\top$  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. In this context, it is also important to note that the response and auxiliary variables are assumed to be error-free and the resulting errors for the point estimates reflect only the uncertainty due to sampling.

## 4 Estimators

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

The one-phase estimator for cluster sampling (SRS) constitutes the *status quo* that is currently applied under the existing one-phase sampling design of the German NFI in order to obtain point and variance estimates for the mean timber volume of a given estimation unit. In order to provide all estimators in the infinite population framework and ensure a consistent terminology with the two-phase estimators in Section 4.2, we will introduce the SRS estimator that is applied in the BWI3 algorithms ([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 to the forest/non-forest decision for each sample location  $x_l$ :

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

The local density on cluster level  $Y_c(x)$ , which is in our case the timber volume per hectare,

#### 4. ESTIMATORS

---

260 is then defined as the average of the individual sample plot densities  $Y(x_l)$ :

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

279 **4.2 Design-based small area regression estimators for cluster  
280 sampling**

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

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

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

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

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

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

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

294 with

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

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

297 In the following, we will give a short description of each small area estimator and refer to  
298 [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.

302

303 **4.2.1 Pseudo Small Area Estimator (PSMALL)**

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

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

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

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

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

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

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

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

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

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

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

321 The estimated design-based variance of  $\hat{Y}_{c,G,PSMALL}$  is given by Eq. 14b. Basically, the  
 322 first term constitutes the variance introduced by the uncertainty in the regression coefficients,

## 4. ESTIMATORS

---

324 whereas the second term expresses the variance caused by estimating the exact auxiliary  
325 mean in  $G$  using a non-exhaustive sample  $s_{1,G}$ . The third term is the variance of the model  
326 residuals and thus accounts for the inaccuracies of the model predictions. Note that the first  
327 term can also be rewritten using g-weights (Mandallaz et al, 2016, pg.14) which ensures some  
328 beneficial calibration of the auxiliary variables to the first-phase sample.

329

### 330 4.2.2 Pseudo Synthetic Estimator (PSYNTH)

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

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

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

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

343

### 344 4.2.3 Extended Pseudo Synthetic Estimator (EXTPSYNTH)

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

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

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

356 However, it is important to note that  $\tilde{R}_{c,G} = 0$  under the extended regression model only  
 357 holds if the sample plots  $x_1, \dots, x_l$  of a cluster are *all* either inside or outside the small area,  
 358 i.e.  $M_G(x) \equiv M(x)$ , and thus  $I_{c,G}(x) = \frac{M_G(x)}{M(x)}$  can only take the values 1 or 0. **Mandallaz**  
 359 **et al (2016)** assumed that the effects on the estimates should be negligible as the number of  
 360 occasions where  $M_G(x) < M(x)$  was considered to be small in practical implementations. It  
 361 was thus a further objective of this study to investigate the actual occurrences and effects of  
 362 this phenomenon by comparing the estimates of EXTPSYNTH to those of PSMALL.

### 363 4.3 Measures of estimation accuracy

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

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

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

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

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

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

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

371 For the PSYNTH estimates, the confidence intervals are

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

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

374 In order to address the potential benefits of the small area estimators compared with the  
 375 SRS approach, we calculated the *relative efficiency* (Eq. 22) which can be interpreted as the

377 relative sample size under SRS needed to achieve the variance under the double-sampling  
378 (DS) estimators.

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

<sup>379</sup> **5 Case study**

<sup>380</sup> **5.1 Study area and small area units**

<sup>381</sup> The German federal state Rhineland-Palatinate (*RLP*) is located in the western part of Germany and borders Luxembourg, France and Belgium. With 42.3% (appr. 8400 km<sup>2</sup>) of the entire state area (19850 km<sup>2</sup>) 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 characterised 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 characterised by large variety of forest structures ranging from characteristic oak coppices (Moselle valley), pure spruce, beech and scots pine forests (i.a. Hunsrück and Palatinate forest) up to mixed forests comprising variable proportions of oak, larch, spruce, Scots pine and beech. Around 82% of the forest area in RLP are mixed forest stands and 69% of the forest area exhibit a multi-layered vertical structure. The forest area of RLP are divided into 3 ownership classes, i.e. state forest (27%), communal forest (46%) and privately owned forest (27%). The forest service of RLP has the legal mandate to sustainably manage the state and communal forest area (73% of the entire forest area), including forest planning, harvesting and the sale of wood ([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 SA units for which design-based small area estimations of the mean standing timber volume were calculated by incorporating the available terrestrial inventory data of the BWI3 in the estimators described in Section 4. The average area of the SA units was 43'777 ha on the FA-level, and 4624 ha on the FR level.

<sup>405</sup> **5.2 Terrestrial sample**

<sup>406</sup> Rhineland-Palatinate (*RLP*) is covered by a 2x2 km inventory grid of the German NFI. In the last inventory (BWI3) conducted in the year 2013, timber volume information was derived for 2810 clusters (8092 plots) in the field survey. The local timber volume density on the plot and cluster level for this sample was consequently calculated according to Section 4.1. In the framework of this survey, the plot center coordinates were re-measured with the differential global satellite navigation system (DGPS) technique. Knowledge about the exact plot positions were considered crucial to provide optimal comparability between the terrestrial observations and the information derived from the auxiliary information. A comparison of the DGPS coordinates with the so-far used target coordinates revealed that 90% of all horizontal deviations lay in the range of 25 meters. A detailed analysis of horizontal DGPS errors in

## 5. CASE STUDY

---

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

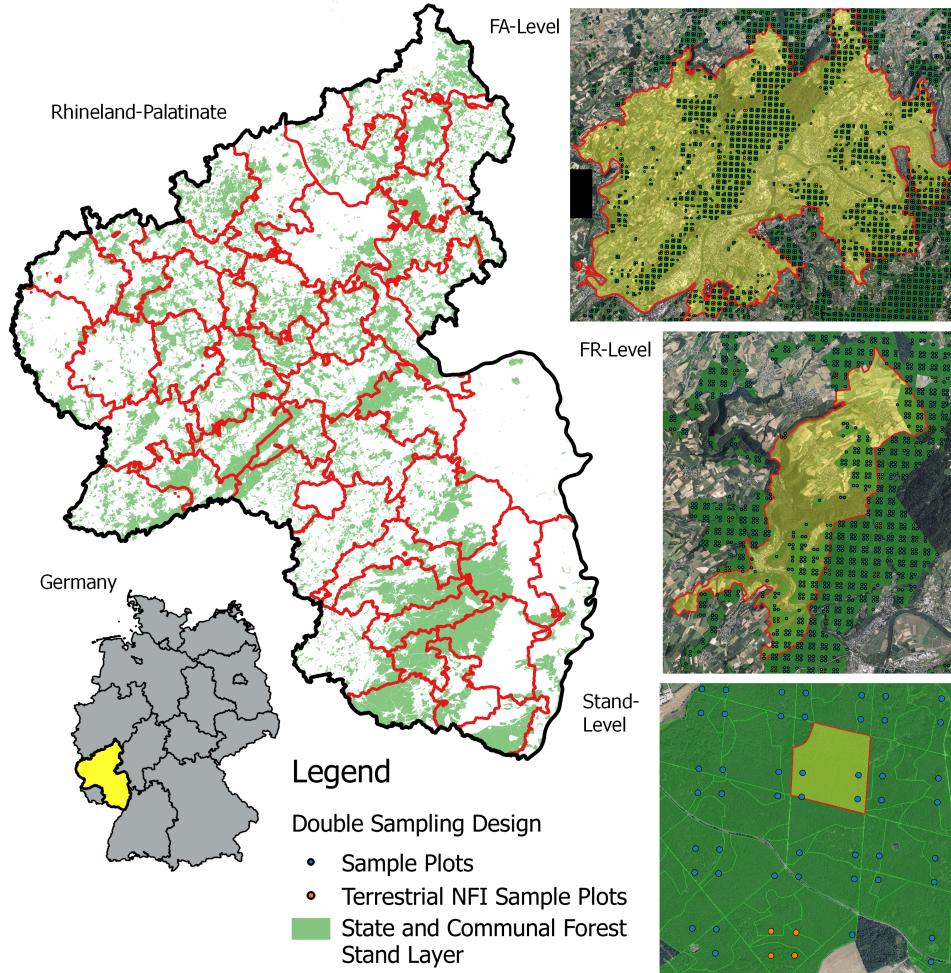


Figure 1: *Left:* Study area with delineated FA forest management units. *Right:* Example for each of the three management units (from top to bottom): FA, FR and forest stand unit overlayed with the extended double-sampling cluster design. *Green:* Forest stand polygon layer defining the forest area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

### 421 5.3 Extension to double-sampling design

422 In order to apply the small area estimators (Section 4.2), the existing NFI design was ex-  
 423 tended to a double-sampling design by densifying the existing systematic 2x2 km grid to a  
 424 grid size of 500x500 m that constituted the large first phase  $s_1$  in accordance to Section 3

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

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

Variable	Mean	SD	Maximum
Timber Volume (m <sup>3</sup> /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

## 437 5.4 Auxiliary data

### 438 5.4.1 LiDAR canopy height model

439 A prerequisite for the application of the suggested two-phase small area estimators is the  
440 identification of suitable auxiliary data available over the entire study area. From 2003 to  
441 2013, the topographic survey institution of RLP conducted an airborne laserscanning acqui-  
442 sition over the entire federal state during leaf-off conditions in order to derive a countrywide  
443 digital terrain model (DTM) as well as a digital surface model (DSM). For this study, the  
444 recorded ALS data was used to create a canopy height model (CHM) in raster format, provid-  
445 ing discrete information about the canopy surface height of the forest area in a spatial resolu-  
446 tion of 5 meters (Fig. 2, top). The CHM was calculated as the difference between the digital  
447 terrain model and the digital surface model that were derived by a Delaunay interpolation of  
448 the ground and first ALS pulses respectively. A more detailed description of the procedure can  
449 be found in Hill et al (2018). The CHM provided the most valuable information to be used  
450 in the OLS regression model for predicting the timber volume on the plot and cluster level.  
451 However, it should be noted that the prolonged acquisition period of the ALS campaign led to  
452 the possibility of poor temporal alignment with the BWI3 survey, sometimes up to 10 years.  
453 In addition, the quality of the CHM varied substantially as ALS technology evolved over the

years. For example, the ALS acquisitions recorded in 2002 and 2003 exhibited particularly poor quality with about only 0.04 point per  $\text{m}^2$ , whereas more recent datasets contained more than 5 points per  $\text{m}^2$ . Furthermore, CHM information was not available at 16 sample locations due to sensor failures. These plots were deleted from the sampling frame and treated as missing at random. This assumption was considered to be reasonable as the respective sample locations did not exclude specific forest structures.

#### 5.4.2 Tree species map

Additional auxiliary data was derived from a countrywide satellite-based classification map predicting the five main tree species (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 5x5 m and was calculated from 22 bi-temporal satellite images (SPOT5 and RapidEye) using a spatially adaptive classification algorithm (Stoffels et al, 2012). As timber volume estimation on the tree level is often based on species-specific biomass and volume equations, the use of tree species information has often been stated as a key factor for improving the precision of timber volume estimates (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  $\text{km}^2$  in the south-west part of RLP covering an entire FA unit consisting of 10 FR units. In 9 additional FR units, the tree species information was also missing for a subset of the sample locations due to two additional patches with areas of 76  $\text{km}^2$  and 100  $\text{km}^2$  respectively in the northern part of RLP. For these 19 FR units, small area estimation was thus restricted to using only the available CHM information in the regression model. Thus, 411 of 5791 sample locations (approximately 7%) used to fit the regression model were affected by missing tree species information. A summary of the sample sizes and missing auxiliary data for both the CHM and the tree species map is provided in Table 2.

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

<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 <i>and</i> TSPEC	3	2	0	0
missing CHM <i>or</i> TSPEC	7075	3595	414	385

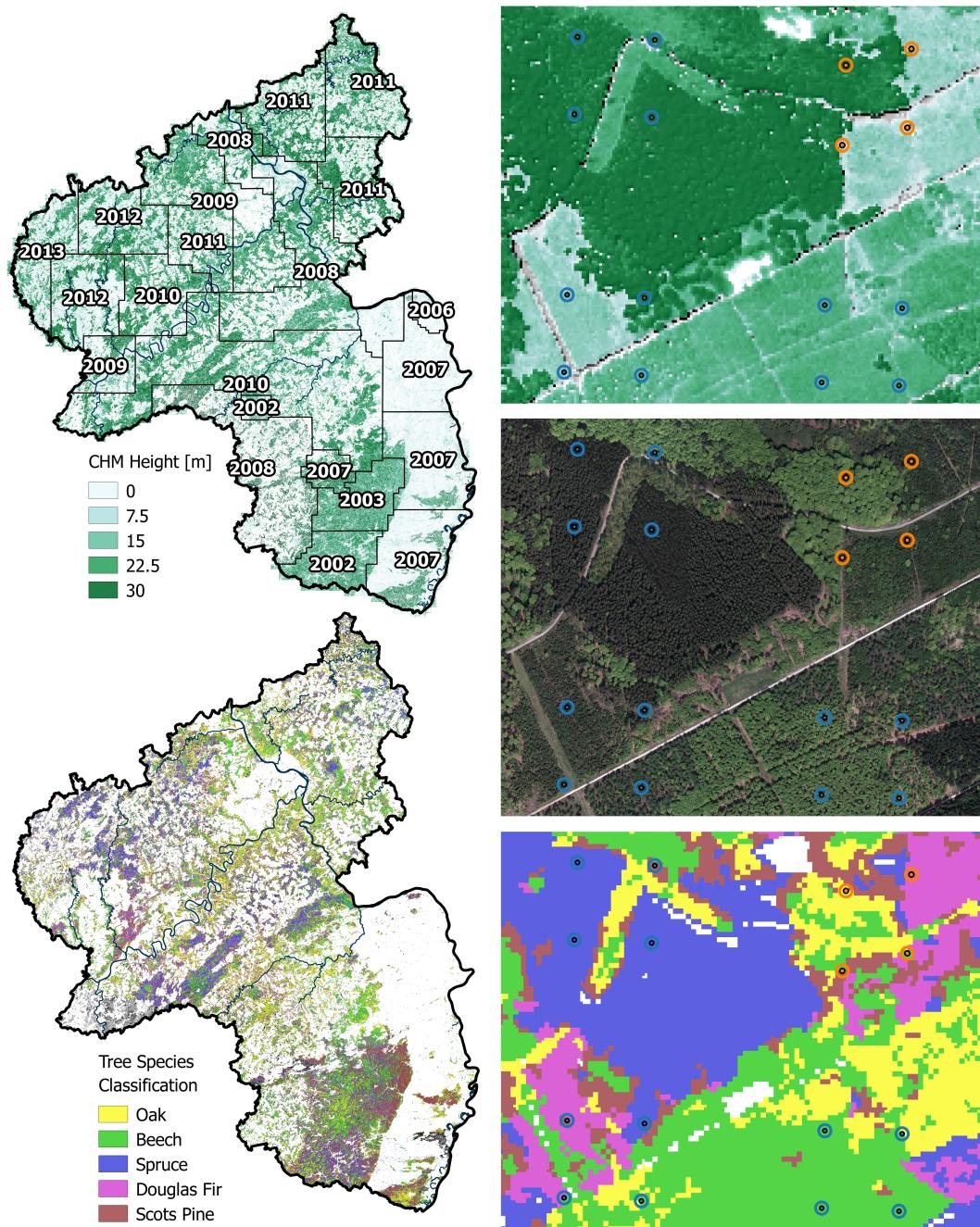


Figure 2: Left: CHM (top) and tree species classification map (bottom) available on the federal state level. Right: Magnified illustration of the supports used to derive the explanatory variables from the auxiliary data.

481 **5.5 Calculation of the explanatory variables**

482 **5.5.1 Canopy height model**

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

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

503 **5.5.2 Tree species map**

504 The tree species map was used to predict the main tree species at each sample plot which  
505 served as an additional categorical variable called *treespecies*. This involved two consecutive  
506 processing steps. In the first step, one of the five tree species was assigned to a sample location  
507 if 100% of the raster values within the edge-corrected support were classified as that species.  
508 Otherwise, the sample location was assigned the value 'mixed'. Likewise for the CHM vari-  
509 ables, the support radius was 12 meters although the use of different support sizes for each  
510 explanatory variable would be in agreement with the two-phase estimators presented in Sec-  
511 tion 4.2. When using the *treespecies* variable in a regression model, the support size and the  
512 percentage threshold parameters had to be optimized in order to minimize the variance within  
513 each level which subsequently leads to improved model precision. A detailed analysis and  
514 description of the optimal parameter processing for the explanatory variables of the present  
515 data set is provided in Hill et al (2018). In a second step, the *treespecies* variable was also  
516 passed through a calibration model in order to reduce the effects of misclassification errors  
517 on the regression model coefficients and to increase model accuracy. The calibration model  
518 consisted of a decision tree from a random forest algorithm (Breiman, 2001) that was trained  
519 to predict the actual main plot tree species (known for all terrestrial plots) based on available

520 auxiliary variables. These variables were the predicted *treespecies* variable, the mean canopy  
 521 height and standard deviation of the CHM, as well as the proportion of coniferous trees esti-  
 522 mated from the classification map and the growing region derived from a polygon map. The  
 523 algorithm was grown with 2000 trees considering 3 of the predictors for each split. We thus  
 524 applied this calibration model to the *treespecies* variable derived at all sample locations  $s_1$ .  
 525 Table 3 gives the classification accuracies (Congalton and Green, 2008) of the *treespecies*  
 526 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[%]	$n_{ref}$	$n_{class}$
Beech	22.31	47.02	883	419
Douglas Fir	24.78	48.72	230	117
Oak	11.07	48.48	289	66
Spruce	53.15	61.13	651	566
Scots Pine	22.91	46.07	179	89
Mixed	84.49	64.53	3152	4127
Overall accuracy: 61.96%			5384	5384

## 527 5.6 Regression Model

528 The model selection process for this study required a substantial time commitment due  
 529 to sophisticated challenges such as: a) the heterogeneity of the remote sensing data, b) the  
 530 identification of the optimal support sizes under angle count sampling, and c) the incorporation  
 531 of tree species information. Here, only a summary of the extensive analysis that was performed  
 532 is provided but the reader can refer to Hill et al (2018) if more details are desired

533 The model with highest adjusted  $R^2$  and lowest RMSE was achieved using *meanheight*,  
 534 *meanheight*<sup>2</sup>, *stddev*, *ALSpyear* and *treespecies* as main effects, and including interaction terms  
 535 between *meanheight* and *ALSpyear*, *stddev* and *ALSpyear*, *meanheight* and *stddev*, and *mean-*  
 536 *height* and *treespecies*. Summary information about the adjusted  $R^2$ , RMSE and RMSE% of  
 537 the selected models is provided in Table 4. The two-phase estimators described in Section  
 538 4.2 derive and apply the regression coefficients and the residuals on the aggregated cluster  
 539 level. We re-evaluated the model as used in the estimators on the cluster level (formulas given  
 540 in Appendix) and found improved model fits compared to the plot level (adjusted  $R^2$  of 0.59  
 541 and RMSE of 101.61 m<sup>3</sup>/ha and 33.6%). The stratification by the ALS acquisition year sub-  
 542 stantially improved the model fit, indicating that it is an effective means in accounting for the  
 543 noise in the data caused by ALS quality variations and time-gaps between the ALS and the  
 544 terrestrial survey. However, the stratification led to a highly unbalanced data set when a further  
 545 *treespecies* stratification was included. For this reason, a individual species modeling within  
 546 each *ALSpyear* stratum remained infeasible, but might have further improved the model fit. An  
 547 additional evaluation of the model's performance within each ALS acquisition year stratum

548 revealed that the quality of the model fit substantially varied between the strata (Table 5). In  
 549 particular, values above the overall adjusted  $R^2$  were higher in ALS acquisition years close to  
 550 the terrestrial survey date compared to years with larger time gaps.

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

Table 4: Model fit 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 <sup>2</sup> +	full model	0.58	90.11	29.76
treespecies + ALSyear +		(0.48)	(139.22)	(45.98)
meanheight:treespecies +				
meanheight:ALSpyear + meanheight:stddev +				
stddev:ALSpyear				
meanheight + stddev + meanheight <sup>2</sup> +	reduced model	0.55	95.23	31.65
ALSpyear + meanheight:ALSpyear +		(0.45)	(144.13)	(47.60)
meanheight:stddev + stddev:ALSpyear				

558 Concerning the existence of outliers or leverage points in the training set for the model,  
 559 it should be noted that it is more problematic for PSMALL, PSYNTH and EXTPSYNTH to  
 560 simply remove them as one might be inclined to do in a model-dependent context. Strictly  
 561 speaking, outlier removal in the design-based context essentially means that those plots, and  
 562 implicitly any potentially similar plots that were not realized in the selected sample, have been  
 563 removed from the sampling frame and are no longer considered part of the forest area of inter-  
 564 est. While this may be valid for some obvious typos or measurement errors, it is generally not  
 565 advisable to manipulate the sampling frame after observing data collected from it, especially  
 566 when the observation in question lies within the small area of interest. However, for sake of  
 567 completeness, we conducted an analysis of influential observations (Fahrmeir et al, 2013, pp.  
 568 160–167) on the plot level for the full regression model. We calculated the leverage values  
 569 and found that 10% of all observations exceeding a predefined critical threshold, i.e. twice the  
 570 average of the hat matrix diagonal entries. Further investigation revealed that several leverage  
 571 points showed unusually large *meanheight* values compared to their respective timber volume  
 572 densities. They tended to occur in ALS acquisition years with longer time gaps to the terres-  
 573 trial survey date and were thus more likely caused by harvesting activities in the sample plot  
 574 area. Although these areas likely affected by harvest should clearly not be removed from the  
 575 sampling frame, it does provide more justification for the inclusion of the *ALSpyear* variable

## 5. CASE STUDY

---

<sup>576</sup> to mitigate these effects.

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

$ALSyyear$	$Area_{ALSyyear}$	$R^2$	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 )

## 577 6 Results

### 578 6.1 General estimation results

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

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

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

### 592 6.2 Estimation error

593 On both small area levels, the design-unbiased estimators PSMALL and EXTPSYNTH led  
 594 to a substantial reduction in the estimation error compared to the SRS estimator (Fig. 3). On  
 595 the FA level, the SRS estimator yielded an estimation error of 6.7% on average compared to  
 596 5.2% and 4.8% under EXTPSYNTH and PSMALL respectively (Table 6). The cumulative  
 597 error distribution (Fig. 3, left) reveals that under the SRS estimator, errors less than 5% were  
 598 achieved for 17% of the FA units (8 of 45). This proportion could be increased to 62% (28

## 6. RESULTS

---

599 FA units) and 73% (33 FA units) by application of the PSMALL and EXTPSYNTH estimator. 600 95% of all estimates exhibited errors less than 9.5% under the SRS estimator and less 601 than 6.6% when using PSMALL or EXTPSYNTH. Estimation errors higher than 10% only 602 appeared twice for each of the three estimators.

603 Although the estimation errors were substantially larger overall on the FR level compared 604 to the FA level due to smaller sample sizes, the error reduction from SRS by PSMALL and 605 EXTPSYNTH were even more pronounced (Fig. 3, right). The average error under the SRS 606 estimator was 18.3%, while it was 11.3% and 12.2% under PSMALL and EXTPSYNTH (Ta- 607 ble 6). Errors smaller than 10% were achieved for 15% of the FR units by the SRS estimator, 608 and for 46% by the PSMALL and PSYNTH estimator. 95% of the 321 FR units where PS- 609 MALL and EXTPSYNTH could be applied exhibited errors less than 20%. In comparion, the 610 SRS estimates resulted in errors less than 36.6% for 95% of the 388 FR units.

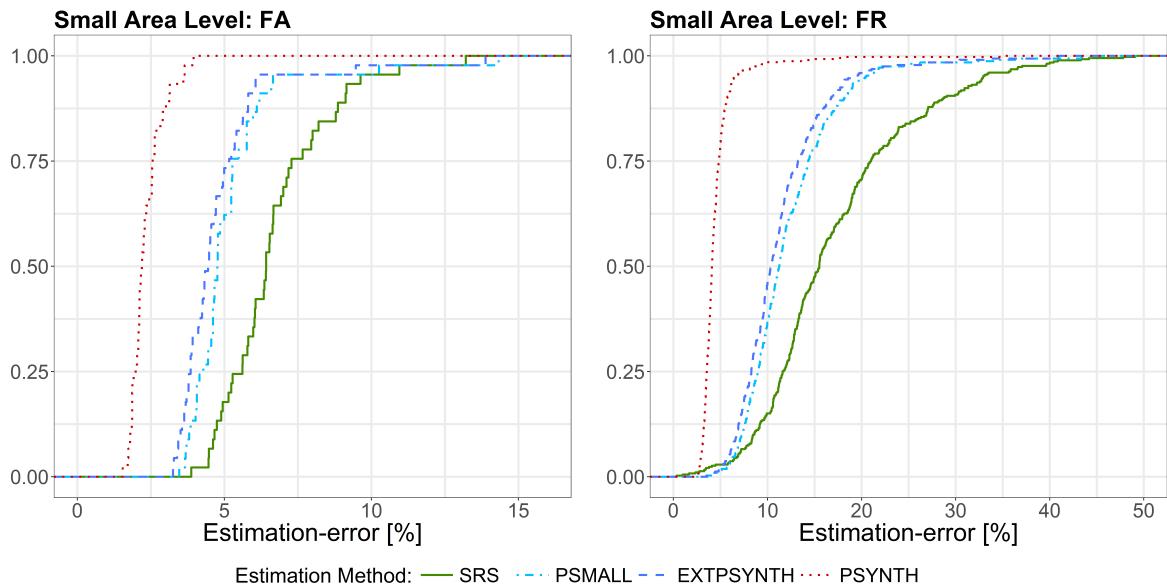


Figure 3: Cumulative distribution of estimation errors under SRS, PSMALL, EXTPSYNTH and the PSYNTH estimator. *Left:* Results for the 45 FA units. *Right:* Results for the 388 (SRS), 321 (PSMALL, EXTPSYNTH) and 403 (PSYNTH) FR units.

611 On both small area levels, the PSYNTH estimator resulted in much smaller estimation er- 612 rors compared to PSMALL and EXTPSYNTH. This was as expected, since the PSYNTH 613 variance estimate does not take the residual variation in each small area unit into account 614 (Section 4.2.2). Compared to the asymptotically design-unbiased estimators PSMALL and 615 EXTPSYNTH, the estimation errors produced by PSYNTH thus seem to be too optimistic. 616 One should also recall that the estimates of the PSYNTH estimator are potentially design- 617 biased.

### 6.3 Comparison of PSMALL and EXTPSYNTH

Figure 3 reveals that the error distribution of PSMALL and EXTPSYNTH are very similar, with PSMALL showing marginally higher estimation errors. In order to investigate the differences between PSMALL and EXTPSYNTH, we compared the g-weight variances of both estimators for all 321 FR units (Fig. 4, left). As obvious, PSMALL yielded slightly larger variances for the vast majority of the estimates. As addressed in Section 4.2.3, one possible explanation for differences was the effect of one or more 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%). We compared the variances of PSMALL and EXTPSYNTH for all small areas that did not have the violations using a Wilcoxon Rank-Sum Test (Wilcoxon et al, 1970). This test was also performed for groups  $n_{2,G} \leq 6$ ,  $n_{2,G} > 6$  and  $n_{2,G} > 10$ . The distribution of variances from EXTPSYNTH were found to be significantly lower than that of PSMALL (p-values of  $1.2 \times 10^{-13}$ ,  $3.4 \times 10^{-8}$ ) except for the group of  $n_{2,G} > 10$  (p-value of 0.125). The latter was expected since the variances of both estimators are asymptotically equivalent under large terrestrial sample sizes  $n_{2,G}$  within the small area (Mandalaz et al, 2016, pp.17–18). This was also confirmed by a comparison of the absolute differences in the variances (Fig. 4, right) which decreased with increasing terrestrial sample size. The Wilcoxon Rank-Sum Test was also used to compare the EXTPSYNTH variances of those small areas with violation (depicted in red diamonds, Fig. 4) to the PSMALL variances without violations. Overall as well as for  $n_{2,G} \leq 6$  there was a significant difference, but not for  $n_{2,G} > 6$ . This provided some evidence that the violations created a statistically significant influence on the EXTPSYNTH variance that makes it appear to be slightly over-optimistic. However, a comparison of the confidence intervals of PSMALL and EXTPSYNTH revealed that the variance differences did not lead to statistically significant point estimates.

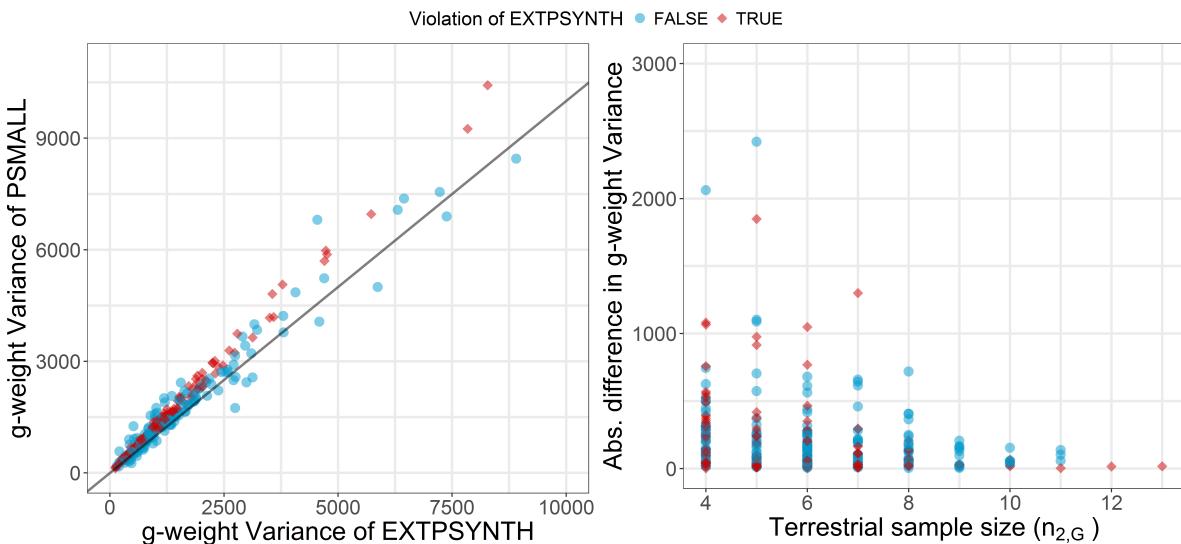


Figure 4: *Left:* Comparison of the g-weight variance between the PSMALL and the EXTPSYNTH estimator for the 321 FR units. *Right:* Difference in g-weight variance between the PSMALL and the EXTPSYNTH estimator in dependence of the terrestrial data ( $n_{2,G}$ ) in the FR unit.

## 6. RESULTS

---

### 6.4 Variance reduction compared to SRS

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

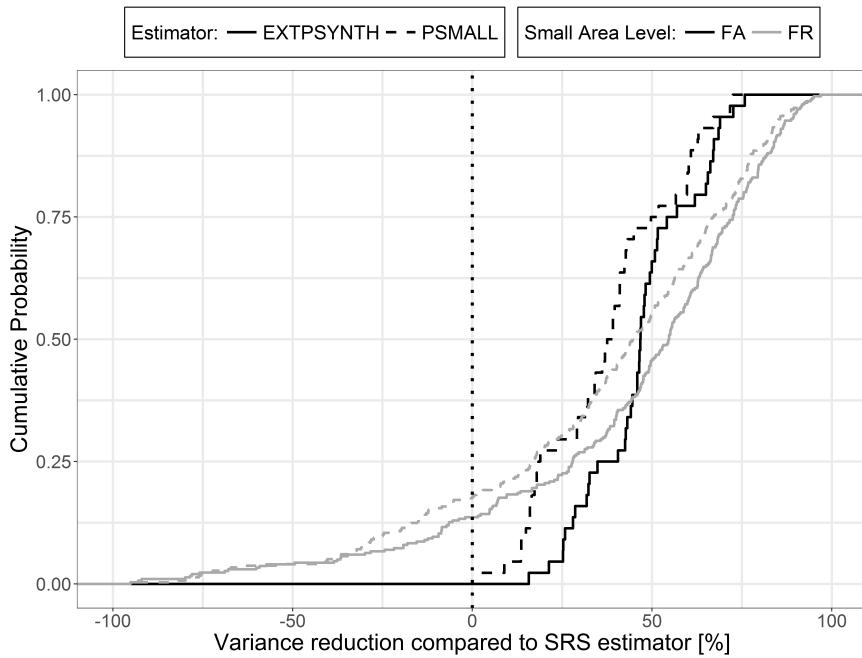


Figure 5: Cumulative distribution of variance reduction by the PSMALL and EXTPSYNTH compared to the SRS estimator for the 45 FA and 321 FR units.

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

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

## 6. RESULTS

---

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

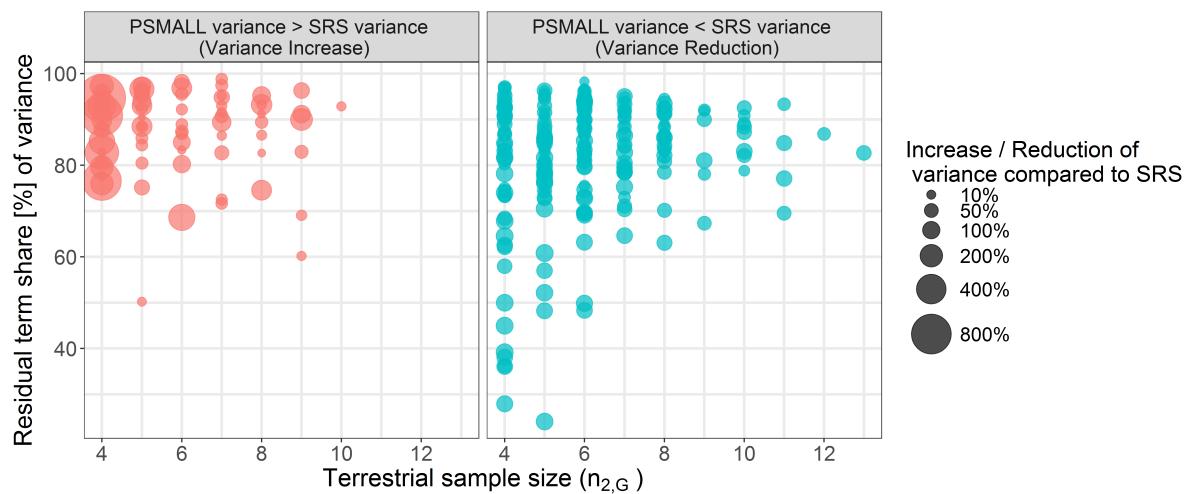


Figure 6: Share of the overall variance by the residual term of the PSMALL estimator for various small area sample sizes. Points are scaled by the overall percentage reduction/increase of the variance compared to SRS.

680 **7 Discussion**

681 **7.1 Performance of estimators**

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

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

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

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

720 trial data in the small area unit, which is primarily a means to ensure the zero mean residual  
721 property of the EXTPSYNTH. However, our analysis indicated that the difference between  
722 the two estimators is negligible for sample sizes  $\geq 10$  due to their asymptotic equivalency.  
723 Furthermore, one or more clusters not entirely included in the small area unit did not have a  
724 notable impact on the estimates of EXTPSYNTH when the terrestrial sample size was more  
725 than 6. However, there was a slight but statistically significant tendency to be over-optimistic  
726 for sample sizes between 4 and 6. More empirical evidence must be gathered before general-  
727 izing this as a rule of thumb for the application of the 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 the substantial introduction of artificial noise in the data.  
737 Whereas a post-stratification to the ALS acquisition years was an effective means to counter-  
738 act the implied residual inflation, several leverage points were unambiguously caused by the  
739 temporal asynchronicity. Undetectable forest loss during the gap between the ALS acquisition  
740 and the NFI was also likely a cause for high residual variance in some small areas compared to  
741 the terrestrial data variance, which subsequently led to higher variances than the SRS estima-  
742 tor. As opposed to the ALS data, the availability of a country-wide tree species classification  
743 map has yet been unique among all German federal states. Whereas the study of [Hill et al](#)  
744 ([2018](#)) already showed that the tree species information was able to improve the model fit,  
745 it has yet not been used to its full potential. One reason for this was the impossibility of  
746 modeling individual tree species within each ALS acquisition year, which would add further  
747 explanatory power. Another reason was the lack of available satellite data for classification  
748 in some parts of the country, which led to missing values in the inventory data and restricted  
749 19 FR units to a simpler regression model. Promising steps with respect to more up-to-date  
750 canopy height information have already been made, as the topographic survey institution of  
751 RLP will from this year on provide a country-wide canopy height model derived from aerial  
752 imagery acquisitions. These campaigns will in the future be conducted in a two-year period  
753 and allow to derive canopy height information matching the dates of terrestrial forest invento-  
754 ries. A study of [Kirchhoefer et al](#) ([2017](#)) recently indicated that similar model performance for  
755 German NFI data can be achieved using such imagery-based canopy height models. Due to the  
756 improved coverage and repetition rate of the Sentinel-2 satellite ([ESA, 2017](#)), the tree species  
757 classification map will in the future be updated each year. We consider these alternative aux-  
758 iiliary data sources to also solve the problem of missing explanatory variables at inventory  
759 plots. One could also make use of the exhaustive information within the two-phase estima-

760 tors by using the true the auxiliary means (Mandallaz, 2013a; Mandallaz et al, 2013), which  
761 could further decrease estimation errors. Previous studies of Mandallaz et al (2013) however  
762 showed that given a reasonable large sample size of the first phase, the differences in the esti-  
763 mation error are usually small. With respect to the substantial improvements in the temporal  
764 synchronicity between auxiliary and terrestrial inventory data, we consider the demonstrated  
765 double-sampling approach also to be very efficient for change estimation (Massey and Man-  
766 dallaz, 2015).

## 767 8 Conclusion

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

## 784 Acknowledgements

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

<sup>793</sup> **A Appendix**

<sup>794</sup> **R-squared on cluster level**

<sup>795</sup> The  $R^2$  on the cluster level is calculated using the number of plots  $M(x)$  of each cluster in  
<sup>796</sup> order to weight for the varying number of plots on which  $Y_c(x)$  and  $\hat{Y}_c(x)$  are based on.

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

<sup>797</sup>  $Y_c(x)$  and  $\hat{Y}_c(x)$  are the predicted and observed local densities on the cluster level calculated ac-  
<sup>798</sup> cording to Equations 2 and 12.  $\hat{\bar{Y}}_c$  is the estimated sample mean corresponding to the weighted  
<sup>799</sup> mean over all observed local densities on the cluster level (Eq. 8).

<sup>800</sup> **RMSE on cluster level**

<sup>801</sup> The same weights  $M(x)$  are also applied to calculate the RMSE on the cluster level.  $n_2$  is  
<sup>802</sup> the number of clusters used in the modeling frame.

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

<sup>803</sup> The *relative* or *normalized* RMSE is calculated by dividing the RMSE by the estimated sample  
<sup>804</sup> mean  $\hat{\bar{Y}}_c$ :

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

<sup>805</sup> Note that the weights  $\frac{M(x)}{\bar{M}_2} \equiv 1$  if the number of plots per cluster is constant.

## 806 References

- 807 Bitterlich W (1984) The relascope idea. Relative measurements in forestry. Commonwealth  
808 Agricultural Bureaux
- 809 Böckmann T, Saborowski J, Dahm S, Nagel J, Spellmann H (1998) A new conception for  
810 forest inventory in lower saxony. Forst und Holz (Germany)
- 811 Breidenbach J, Astrup R (2012) Small area estimation of forest attributes in the norwegian  
812 national forest inventory. European Journal of Forest Research 131(4):1255–1267, DOI  
813 10.1007/s10342-012-0596-7
- 814 Breiman L (2001) Random forests. Machine learning 45(1):5–32, DOI 10.1023/A:  
815 1010933404324
- 816 Congalton RG, Green K (2008) Assessing the accuracy of remotely sensed data: principles and  
817 practices. CRC press, DOI 10.1201/9781420055139, URL <https://doi.org/10.1201/9781420055139>
- 819 Bundesministerium für Ernährung LuV (2011) Aufnahmeanweisung für die dritte Bun-  
820 deswaldinventur BWI3 (2011 - 2012). URL <https://www.bundeswaldinventur.de/index.php?id=421>
- 822 ESA (2017) Sentinel-2 earth observation mission. URL [http://www.esa.int/Our\\_Activities/Observing\\_the\\_Earth/Copernicus/Sentinel-2](http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2), accessed: 2017-03-29
- 824 Fahrmeir L, Kneib T, Lang S, Marx B (2013) Regression: models, methods and applica-  
825 tions. Springer Science & Business Media, DOI 10.1007/978-3-642-34333-9, URL <http://www.springer.com/us/book/9783642343322>
- 827 Gauer J, Aldinger E (2005) Waldökologische naturräume deutschlands-wuchsgebiete. Mit-  
828 teilungen des Vereins für Forstliche Standortskunde und Forstpflanzenzüchtung 43:281–288
- 829 Goerndt ME, Monleon VJ, Temesgen H (2011) A comparison of small-area estimation tech-  
830 niques to estimate selected stand attributes using lidar-derived auxiliary variables. Canadian  
831 journal of forest research 41(6):1189–1201
- 832 Grafström A, Schnell S, Saarela S, Hubbell S, Condit R (2017) The continuous population  
833 approach to forest inventories and use of information in the design. Environmetrics
- 834 Gregoire TG, Valentine HT (2007) Sampling strategies for natural resources and the environ-  
835 ment. CRC Press
- 836 Hill A, Massey A (2017) forestinventory: Design-Based Global and Small-Area Estima-  
837 tions for Multiphase Forest Inventories. R package version 0.3.1. URL <https://cran.r-project.org/web/packages/forestinventory/index.html>

- 839 Hill A, Buddenbaum H, Mandallaz D (2018) Combining canopy height and tree species map  
840 information for large scale timber volume estimations under strong heterogeneity of auxil-  
841 iary data and variable sample plot sizes, submitted for publication
- 842 Kirchhoefer M, Schumacher J, Adler P, Kändler G (2017) Considerations towards a novel ap-  
843 proach for integrating angle-count sampling data in remote sensing based forest inventories.  
844 *Forests* 8(7):239
- 845 Koch B (2010) Status and future of laser scanning, synthetic aperture radar and hyperspec-  
846 tral remote sensing data for forest biomass assessment. *ISPRS Journal of Photogram-  
847 metry and Remote Sensing* 65(6):581–590, DOI 10.1016/j.isprsjprs.2010.09.001, URL  
848 <https://doi.org/10.1016/j.isprsjprs.2010.09.001>
- 849 Köhl M, Magnussen SS, Marchetti M (2006) Sampling methods, remote sensing and GIS  
850 multiresource forest inventory. Springer Science & Business Media
- 851 Kublin E, Breidenbach J, Kändler G (2013) A flexible stem taper and volume prediction  
852 method based on mixed-effects b-spline regression. *European journal of forest research*  
853 132(5-6):983–997, DOI 10.1007/s10342-013-0715-0
- 854 Kuliešis A, Tomter SM, Vidal C, Lanz A (2016) Estimates of stem wood increments in forest  
855 resources: comparison of different approaches in forest inventory: consequences for inter-  
856 national reporting: case studyof european forests. *Annals of Forest Science* 73(4):857–869
- 857 Lamprecht S, Hill A, Stoffels J, Udelhoven T (2017) A machine learning method for co-  
858 registration and individual tree matching of forest inventory and airborne laser scan-  
859 ning data. *Remote Sensing* 9(5), DOI 10.3390/rs9050505, URL <http://www.mdpi.com/2072-4292/9/5/505>
- 861 von Lüpke N (2013) Approaches for the optimisation of double sampling for stratification in  
862 repeated forest inventories. PhD thesis, University of Göttingen
- 863 von Lüpke N, Hansen J, Saborowski J (2012) A three-phase sampling procedure for contin-  
864 uous forest inventory with partial re-measurement and updating of terrestrial sample plots.  
865 *European Journal of Forest Research* 131(6):1979–1990, DOI 10.1007/s10342-012-0648-z
- 866 LWaldG (2000) Landeswaldgesetz Rheinland-Pfalz (Forest Act Rhineland-Palatinate). URL  
867 <http://www.landesrecht.rlp.de>, Rhineland-Palatinate, Germany
- 868 Magnussen S, Mandallaz D, Breidenbach J, Lanz A, Ginzler C (2014) National forest inven-  
869 tories in the service of small area estimation of stem volume. *Canadian Journal of Forest  
870 Research* 44(9):1079–1090
- 871 Magnussen S, Mauro F, Breidenbach J, Lanz A, Kändler G (2017) Area-level analysis of forest  
872 inventory variables. *European Journal of Forest Research* pp 1–17
- 873 Mandallaz D (2008) Sampling techniques for forest inventories. CRC Press, DOI 10.1201/  
874 9781584889779, URL <https://doi.org/10.1201/9781584889779>

- 875 Mandallaz D (2013a) Design-based properties of some small-area estimators in forest inven-  
876 tory with two-phase sampling. Canadian Journal of Forest Research 43(5):441–449, DOI  
877 10.1139/cjfr-2012-0381, URL <https://doi.org/10.1139/cjfr-2012-0381>
- 878 Mandallaz D (2013b) A three-phase sampling extension of the generalized regression estima-  
879 tor with partially exhaustive information. Canadian Journal of Forest Research 44(4):383–  
880 388, DOI 10.1139/cjfr-2013-0449, URL <https://doi.org/10.1139/cjfr-2013-0449>
- 881 Mandallaz D, Breschan J, Hill A (2013) New regression estimators in forest inventories  
882 with two-phase sampling and partially exhaustive information: a design-based monte  
883 carlo approach with applications to small-area estimation. Canadian Journal of Forest  
884 Research 43(11):1023–1031, DOI 10.1139/cjfr-2013-0181, URL [https://doi.org/10.](https://doi.org/10.1139/cjfr-2013-0181)  
885 <https://doi.org/10.1139/cjfr-2013-0181>
- 886 Mandallaz D, Hill A, Massey A (2016) Design-based properties of some small-area estima-  
887 tors in forest inventory with two-phase sampling - revised version. Tech. rep., Department  
888 of Environmental Systems Science, ETH Zurich, DOI 10.3929/ethz-a-010579388, URL  
889 <https://doi.org/10.3929/ethz-a-010579388>
- 890 Massey A, Mandallaz D (2015) Design-based regression estimation of net change for for-  
891 est inventories. Canadian Journal of Forest Research 45(12):1775–1784, DOI 10.1139/  
892 <https://doi.org/10.1139/cjfr-2015-0266>, [https://doi.](https://doi.org/10.1139/cjfr-2015-0266)  
893 <https://doi.org/10.1139/cjfr-2015-0266>
- 894 Massey A, Mandallaz D, Lanz A (2014) Integrating remote sensing and past inventory data  
895 under the new annual design of the swiss national forest inventory using three-phase design-  
896 based regression estimation. Canadian Journal of Forest Research 44(10):1177–1186, DOI  
897 10.1139/cjfr-2014-0152, URL <https://doi.org/10.1139/cjfr-2014-0152>
- 898 Naesset E (2014) Area-based inventory in norway – from innovation to an operational reality.  
899 In: Forest Applications of Airborne Laser Scanning - Concepts and Case Studies, Springer,  
900 chap 11, pp 216–240, DOI 10.1007/978-94-017-8663-8
- 901 Polley H, Schmitz F, Hennig P, Kroher F (2010) Germany. In: National Forest Invento-  
902 ries - Pathways for Common Reporting, Springer, chap 13, pp 223–243, DOI 10.1007/  
903 978-90-481-3233-1
- 904 Rao JN (2015) Small-Area Estimation. Wiley Online Library
- 905 Saborowski J, Marx A, Nagel J, Böckmann T (2010) Double sampling for stratification  
906 in periodic inventories—finite population approach. Forest ecology and management  
907 260(10):1886–1895
- 908 Särndal CE, Swensson B, Wretman J (2003) Model assisted survey sampling. Springer Sci-  
909 ence & Business Media

- 910 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)
- 913 Steinmann K, Mandallaz D, Ginzler C, Lanz A (2013) Small area estimations of proportion  
914 of forest and timber volume combining lidar data and stereo aerial images with terrestrial  
915 data. Scandinavian journal of forest research 28(4):373–385
- 916 Stoffels J, Mader S, Hill J, Werner W, Ontrup G (2012) Satellite-based stand-wise forest cover  
917 type mapping using a spatially adaptive classification approach. European journal of forest  
918 research 131(4):1071–1089
- 919 Stoffels J, Hill J, Sachtleber T, Mader S, Buddenbaum H, Stern O, Langshausen J, Dietz  
920 J, Ontrup G (2015) Satellite-based derivation of high-resolution forest information layers  
921 for operational forest management. Forests 6(6):1982–2013, DOI 10.3390/f6061982, URL  
922 <http://www.mdpi.com/1999-4907/6/6/1982>
- 923 Thünen-Institut (2014) Dritte Bundeswaldinventur 2012. URL <https://bwi.info>, accessed:  
924 2017-02-03
- 925 White JC, Coops NC, Wulder MA, Vastaranta M, Hilker T, Tompalski P (2016) Remote  
926 sensing technologies for enhancing forest inventories: A review. Canadian Journal of Remote  
927 Sensing 42(5):619–641, DOI 10.1080/07038992.2016.1207484, URL <http://dx.doi.org/10.1080/07038992.2016.1207484>
- 929 Wilcoxon F, Katti S, Wilcox RA (1970) Critical values and probability levels for the wilcoxon  
930 rank sum test and the wilcoxon signed rank test. Selected tables in mathematical statistics  
931 1:171–259