

Review - revision

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

Andreas Hill, Daniel Mandallaz, Joachim Langshausen

24. Juni 2018

Reviewer 1:

I have only a single marginal question, and probably I am wrong: $Z_c(x)$ is introduced as a row vector in line 183 and Eqs. 9a and 9b, but it is used as column vector in Eqs. 12, 13 and 14. I would appreciate much if the authors could check whether this inconsistency really exists with respect to notation of $Z_c(x)$.

Answer:

Vectors are always column vectors, i.e. $Z(x)$ and $Z_c(x)$ are column vectors. When they are used for predictions they have to be transposed in order to get a scalar product written as a matrix multiplication, i.e. $\mathbf{Z}_c^\top(x)\hat{\beta}_{c,s_2}$. We added the information that $Z_c(x)$ is a column vector in line 183 to make this more clear for the reader.

Reviewer 2:

General comment on publishing the article in Remote Sensing:

We decided to hand in our article to be published in the *Remote Sensing* special issue *Remote sensing based Forest Inventories from Landscape to Global Scale*, since multiple of the mentioned topics for articles of the special issue are covered by our article (such as 'Application of remote-sensing techniques for monitoring forest attributes at the forest enterprise level', 'Application of remote-sensing techniques for monitoring forest attributes for national forest inventories', 'Large scale in-situ data requirement and sampling design for monitoring forest ecosystems with remote-sensing techniques'). Additionally, of course, not all readers of *Remote Sensing* are specialists in forest inventory, but most of the latter frequently read *Remote Sensing*.

1. L28: "entire forest state area of Germany" I understand what you mean, but it may be misleading as readers might think about the area owned by the German state forest enterprise (Bundesforst). I would suggest to rephrase "forested area of Germany".

Has been changed accordingly.

2. L29: 114'191 ha – A wrong unit was used here, its 114 191 km² (or 11 419 124 ha)..

Has been corrected accordingly.

3. L36: "standwise forest inventories (SFI)"?

Has been changed accordingly.

4. L38-41: Please add the information that the FDI of Lower-Saxony (Betriebsinventur, BI) is only done for forests owned by the state forest enterprise (Lower Saxony State Forests = Niedersächsische Landesforsten, NLF)

The information has been added.

5. L41 As far as I remember, [4] is an article written in German. If so, please indicate in the reference list – it is annoying for readers to search for papers that turn out to be written in a language they possibly do not understand.

We now give the original title in German.

6. L42: As far as I remember, v.Lüpke's dissertation [5] is a cumulative work. If so, please cite the original articles and not the compilation.

Has been changed accordingly.

7. L43: I would suggest to cite the classic textbooks of Cochran (1977) and de Vries (1986) beside the modern literature [6-9]

The respective citations have been added.

8. L43-48: I would suggest to add some words (here or in the discussion) addressing the problems arising with multi-purpose forest inventories (e.g. Corona and Marchetti 2007): The stratification is optimized for a single target variable (in most cases the volume of the growing stock) and the introduction of additional target variables becomes challenging (e.g. Ritter and Saborowski, 2014).

We think that the reviewer might mix up *stratification* (where the sample sizes are optimized for strata prior to the actual conduction of the inventory based on auxiliary data) with *post-stratification* (where the categorical auxiliary data are used in the regression model after the inventory has been conducted). The latter, which is the case applied in our article, does not impose any of the mentioned problems of multi-objective optimization. We thus think that including this topic does not fit well into our article since it is not affecting the methods that we are demonstrating.

-
9. *L46-47: Double-sample procedures have not only been used in the FDI of Lower Saxony, but also in the NFIs of Canada (Gillis et al., 2010) and Switzerland (Lanz et al. 2010), and in different other inventories (e.g. Chojnacky, 1998 or Gasparini et al., 2010).*
We rephrased the sentence and added the suggested references.
 10. *L99 "The German NFI (Bundeswaldinventur, BWI)" I think, the abbreviation introduced in the next line becomes more meaningful when giving the German name here.*
Has been changed.
 11. *L100 "The third and most recent"*
Has been changed.
 12. *L101 '4x4km' I would prefer '4km ' 4km'*
We would like to stick to the given notation in accordance with our previously published article (Hill et. al 2018: Combining canopy height and tree species map information for large-scale timber volume estimations under strong heterogeneity of auxiliary data and variable sample plot sizes, *European Journal of Forest Research*).
 13. *L102 '2x2km' same as above (2km ' 2km)*
Same as above.
 14. *L107: A short explanation of the angle count sampling concept would be helpful for readers who aren't foresters, maybe here or in a corresponding appendix.*
We added an explanation of angle count sampling in Section 2.
 15. *L110/111 "DBH" and "DBH" Please decide whether or not to use italics. Please also add the information that this refers to a height of 1.3 m, Remote Sensing has a lot of readers who are not foresters.*
We added the information about the height. Concerning italics, we decided to use italics when first introducing a notation/term and use normal formatting when using the term afterwards.
 16. *L196 Please also provide reference to the R software itself, and not only to the package.*
The reference has been added.
 17. *L273: I think this is the only occasion you use italics for "RLP".*
We use italics when introducing a term for the first time and use normal formatting when using the term afterwards.
 18. *L283: "communal forests" is most often used as a synonym for "community forests" in context of rural development policy. I think here it is rather an odd translation of the German word "Kommunalwald". Maybe, "municipal forests" is the better translation.*
We changed 'communal' to 'municipal' throughout the entire article.
 19. *L295 "RLP"was already introduced as an abbreviation*
'RLP' has been deleted.
 20. *L287: Instead of "45 Forstämter (FA)" I would use "45 forest districts (Forstämter, FA)"*
Has been changed, and likewise for the Forstreviere which we now call sub-districts.
 21. *L289-290: If the SFI is conducted in a five year cycle, shouldn't the management strategy be also set up for five years, instead of ten years?*
The cycle of the stand wise inventories actually varies between 5 and 10 years and can in some cases even exceed 10 years. The stand wise inventories and the date of their conduction are planned by the state forest service. The management strategy for the stands is, however, always made for the upcoming 10 years ('mittelfristige Planung').
 22. *L295 "2x2km"(see above)*
We would like to keep the '2x2km' notation.
 23. *Figure 1: "communal forests"(see above)*
Has been corrected.
 24. *Table 1: Please remove the digits (or at least reduce the number of digits). Giving values with an "accuracy" of 1/100th of a mm (i.e. 10 mm) for DBH, as an example, really does not make much sense.*
We limited the number of digits for timber volume, DBH and height to one digit and removed the digits for stem number.
 25. *Table 3: I think one digit for the percentage values should be enough.*
Has been changed accordingly.

-
26. Table 6: Please indicate for what variable point estimates are. I suppose it is volume of growing stock per area unit [$\text{m}^3 \text{ha}^{-1}$]? And again, check the number of digits.
Yes! We added the missing information accordingly.
27. L454-458 I understand your intention and the necessity of removing FR-units. However, when comparing the performance of the estimators you should use the same population for all estimators. Removing the "difficult" FR-units (i.e. those with a low sampling density) only for PSMALL and EXTPSYNTH is somewhat unfair against the other estimators, and may bias the results. So, my advice is to recalculate SRS and PSYNTH for the same 321 FRs that were used for PSMALL and EXTPSYNTH. Edit: I noticed you confined your analyses to those 321 FRs later on (Fig 5 and Tab. 7), another reason to also do it here.
We changed the evaluation according to your suggestion and added a sentence in the paragraph above the table. The descriptive summary statistics for the one-phase SRS and the PSYNTH estimator did only differ marginally, with exception of the maximal errors which logically turned out to be lower. We also adjusted the numbers given in Section 6.2 and the graphic of Figure 3 respectively.
28. L517: Please use the same scaling for both x- and y-axis on the left plot. It is irritating that the 1:1 line does not cross the grid-corners.
Has been changed accordingly.
29. L655-656: "Northwest German Forest Research Institution" It should be "Institute" instead of "Institution"
Has been corrected.

Reviewer 3:

- Under the design-based inferential framework, invoking the asymptotical properties of a small-area estimator is a contradiction in terms; the asymptotic properties are based on large-sample assumptions, while SAE estimators are used when the sample size is (too) small. If the sample size for the entire population is very large, then the small-area estimation becomes a common domain estimation problem, and there is no need for tailored SA estimators because the direct estimators are probably the most efficient - see Estevao and Särndal (2004) 'Borrowing strength is not the best technique within a wide class of design-consistent domain estimators. *Journal of Official Statistics*, 20, 645-669' for details. This is an important aspect that it's been often ignored in the model-assisted SA studies. The authors put a great deal of effort in assessing the results by the realized sample sizes in various SAs, however, such comparisons have limited relevance if the asymptotic results cannot be invoked. This is not criticism, personally I do not have a solution, and there is probably no analytical solution to derive the small-sample properties of these estimators. However, I think that rising this issue would add value to the manuscript, by informing the readers of potential drawbacks of the estimation methods.
Asymptotic results for the variances are, given the complexity of the problem, the only available ones. When are asymptotic results approximately valid? If one takes for instance $n=6$ (not a very large number), then the distribution of the mean of $n=6$ uniformly distributed random variables is practically undistinguishable from the normal distribution, so that in this case $n=6$ is the asymptotic validity range. On the other hand with extreme values statistics, $n=100$ can be far too small to reach e.g. the Gumbel distribution. Transferring this to the case of forest inventory, it is actually impossible to define a minimal sample size that always ensures the validity of the central limit theorem for the particular case. However, simulations performed for an artificial example in Mandallaz et al. (2013). New regression estimators in forest inventories with two-phase sampling and partially exhaustive information: a design-based monte carlo approach with applications to small-area estimation. *Canadian Journal of Forest Research*, 43(11), 1023-1031, suggested that for SA problems, a sample size in a small area of $n > 6$ is acceptable to reach the nominal coverage rates of the confidence intervals. Whereas these simulations were performed for simple sampling, re-evaluating the simulation example recently confirmed the same results for cluster sampling. We added a respective remark in the discussion, Section 7.1, line 638-645.
- The authors argue that some PSYNTH and the EXTPSYNTH estimators may not always be design-unbiased for the SAs that do not contain field sampling units. I assume that the authors are well aware that the sample sizes within the SAs are random variables. Under SAE there is always a risk that some SAs will not be represented in a particular sample, but every SA will contain field observations under unconditional estimation over all possible samples, otherwise a coverage bias will occur. Thus, for statements such as 'The PSYNTH estimator thus has a potential unobservable design-based bias' (lines 228-229), the authors should clarify if they are considering the properties of the estimators conditioning on the realized sample, or unconditionally. One could also argue that, if the asymptotic arguments are invoked (although they should probably not be), then the PSYNTH estimator is still design-consistent for SAs, even in the absence of field

observations. See Firth and Bennett (1998) 'Robust models in probability sampling. *Journal of the Royal Statistical Society*, 60, 3-21'.

It is, of course, conditioned on the realized sample. We added this information in the respective line 248/249.

3. Regarding the case study, I understand the motivation for using SAE at FR level, where the average number of field sampling units (clusters) is about 5 (0 to 13) per SA. However, using SAE at FA level - where the average sample size contains about 46 clusters (11 to 64)- has to be better justified because, least for some FAs, a direct estimator would have been more efficient (Estevao and Særdal 2004). However, it is perfectly understandable that constructing domain specific (FA-level) models is a time consuming endeavor and simplifying the regression modeling is a pragmatic solution for statistical production, but a discussion on this topic should be provided to inform the readers of the possible choices.

Thank you for this valuable hint! We added a discussion of this question in the discussion, Section 7.1., line 590-600. We point out that the main reason we did not consider direct estimation on the FA-level was the large number of parameters (39 coefficients for the full model, see also Hill et al. 2018*) to be fitted for each FA-unit model due to the pronounced heterogeneity of the auxiliary data (multiple tree species and ALS acquisition years per FA-unit). In this case, the strategy to 'borrow strength' from the entire inventory domain was preferred to direct estimation in order to avoid overfitting and the implied risk of unstable global estimation.

* Hill, A., Buddenbaum, H., & Mandallaz, D. (2018). Combining canopy height and tree species map information for large-scale timber volume estimations under strong heterogeneity of auxiliary data and variable sample plot sizes. *European Journal of Forest Research*, 1-17.

4. Regarding the terminology used (especially) in Section 4, expressions like "design-based small area regression estimator" are not very accurate, since the same estimator (i.e., the same mathematical expression) can be used (appropriately or not) in different contexts. For instance, one could use the PSYNTH estimator (based on an internal model) as an external estimator for another population in a model-dependent framework. Instead, I would rather use the formulation "design-based small area regression estimation", to avoid confusions.

When using the terms '... estimators' as titles in Section 4, they refer explicitly to the estimators which we decided to apply in our article. In accordance to the various articles published on the used estimators by Mandallaz (see selected references below), we would prefer to maintain the terminology as given in our article. Additionally, up to our knowledge the terms 'design-based estimator', 'model-based estimators' are also frequently used in the literature, e.g.:

- Mandallaz, D. (2013). A three-phase sampling extension of the generalized regression estimator with partially exhaustive information. *Canadian Journal of Forest Research*, 44(4), 383-388.
- Mandallaz, D., Breschan, J., & Hill, A. (2013). New regression estimators in forest inventories with two-phase sampling and partially exhaustive information: a design-based monte carlo approach with applications to small-area estimation. *Canadian Journal of Forest Research*, 43(11), 1023-1031.
- Massey, A., & Mandallaz, D. (2015). Comparison of classical, kernel-based, and nearest neighbors regression estimators using the design-based Monte Carlo approach for two-phase forest inventories. *Canadian Journal of Forest Research*, 45(11), 1480-1488.
- Breidenbach, J., McRoberts, R. E., & Astrup, R. (2016). Empirical coverage of model-based variance estimators for remote sensing assisted estimation of stand-level timber volume. *Remote Sensing of Environment*, 173, 274-281
- Ghosh, M., & Kumar Sinha, B. (1990). On the consistency between model-and design-based estimators in survey sampling. *Communications in Statistics-Theory and Methods*, 19(2), 689-702
- Magnussen, S. (2015). Arguments for a model-dependent inference?. *Forestry: An International Journal of Forest Research*, 88(3), 317-325

and many more.

5. Please elaborate on the importance of the zero-mean residual property (lines 191-192). This is an important issue that can be tracked down to the definition of the internally bias-calibrated models in Firth and Bennett (1998) "Robust models in probability sampling. *Journal of the Royal Statistical Society*, 60, 3-21".

We rephrased this part. We added some sentences to emphasize the difference between the OLS residuals being zero on average (this is of course the case for any OLS fitted model) and the consequence for the theoretical residuals in the inventory domain F if they model is fitted internally (then, we have the *zero mean residual property* according to the works of Mandallaz). We also mention the importance of this

property for deriving the g-weight variance in the framework of small area estimation, a technique that was introduced by Mandallaz in 2012/2013. In case a reader is interested in more mathematical details, we give the respective references. We would like to emphasize that the objective of our article was to illustrate the application of the already extensively mathematically described estimators by Mandallaz to large inventory areas and investigate their suitability for future operational use in NFIs. Consequently, the article explicitly targets at an audience of practitioners. For this reason, we strongly believe that discussing more mathematical details which have already been given in various peer-reviewed articles in the Canadian Journal of forest research and technical reports (references given in the article) is beyond the scope of the article's objective.

6. *Line 231-232: See the general comments with regard to design-unbiasedness.*

We added that it is conditional on the realized sample.

7. *Line 233: The PSYNTH estimator is not design-based, the estimation is design based.*

Has been changed respectively.

8. *Line 418-419: Is it real stratification of post-stratification by the ALS acquisition year?*

We referred to using a categorical variable in a regression model in a classical model-dependent framework as 'stratification'. Using this model in a regression estimator is indeed referred to as 'post-stratification'. The entire paragraph which the reviewer refers to, however, exclusively deals with the model building part. In order to avoid misunderstandings, we rephrased the sentences and avoided the term 'stratification'.

9. *Lines 435-442: Removing observation during the regression model process is perfectly valid, as long as the regression residuals are estimated on the full sample. Of course, the zero-mean residual property doesn't hold anymore, which in turn restricts the use of some estimators. Please reformulate.*

You are absolutely right. We rephrased the paragraph respectively.

10. *Line 593-594: If the asymptotic argument is raised, then there is no need for small-area estimators, I would say, because the sample would sufficiently large to support a direct estimators.*

We think the reviewer misunderstood the sentence. We here refer to the asymptotic equivalency of the two design-unbiased small area estimators PSMALL and EXTPSYNTH (firstly mentioned in Mandallaz, D.; Hill, A.; Massey, A. (2016) Design-based properties of some small-area estimators in forest inventory with two-phase sampling - revised version. *Technical report*, Department of Environmental Systems Science, ETH Zurich, and empirically observed in our study).