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 3:

1. *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 to 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 where 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.

2. *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 a 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 Saerndal 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).