English language and style

( ) Extensive editing of English language and style required
( ) Moderate English changes required
(x) English language and style are fine/minor spell check required
( ) I don't feel qualified to judge about the English language and style

	Yes	Can be improved	Must be improved	Not applicable
Does the introduction provide sufficient background and include all relevant references?	(x)	( )	( )	( )
Is the research design appropriate?	(x)	( )	( )	( )
Are the methods adequately described?	(x)	( )	( )	( )
Are the results clearly presented?	(x)	( )	( )	( )
Are the conclusions supported by the results?	(x)	( )	( )	( )

Comments and Suggestions for Authors

In "A double-sampling extension of the German National Forest Inventory for design-based area estimation on forest district levels" the authors present a comprehensive application a evaluation of three design-based small area estimators, which are used to derive district-lev (Forstamt, Forstrevier) estimates of the growing stock timber volume based on German nat forest inventory plots in the entire German federal state of Rhineland-Palatinate. The small estimators were formulated in context of a double sampling scheme, and their performance compared with the design-based estimators traditionally used in context of one-phase simp random sampling.

The fundamental inferential framework is based on former work by the co-author D. Mandal is a leading expert in the field of sampling theory for forest inventory. Although the methodo heavily mathematical it is described very clearly and comprehensibly.

During the reading of the first lines of the manuscript, a few questions came into my mind, did the authors use a constant radius of the circular "support area" with 12m, instead of usi individual radius equal to max(dbh)/4 (inclusion zone radius of the thickest tree)? How was missing tree species information; did the authors use extra models with extra parameter es What was the influence of tree species misclassification on the precision of the estimates? effect would have violations of assumptions for extended synthetic estimator (the fraction o plots per cluster is no longer an integer)? To what extend did the temporal misalignment of field measurements and the remote sensing data influence the precision of estimates?

When reading a few pages further, I was very happy to see that the authors did a fantastic the above mentioned questions were carefully and exhaustively analyzed.

In summary, the work represents outstanding research and I am sure that the paper can ev as a practical manual to implement all the novel methodology in the upcoming cycle of the national forest inventory. Furthermore, the presented approaches have also high relevance f world-wide existing forest inventory schemes.

I have only a single marginal question, and probably I am wrong: Z_c(x) is introduced as a in line 183 and Eqs. 9a & 9b, but it is used as column vector in Eqs. 12, 13 & 14. I would a much if the authors could check whether this inconsistency really exists with respect to no Z_c(x).

Finally, I strongly recommend accepting this manuscript for publication after a quick check dimension of Z_c(x).

English language and style

( ) Extensive editing of English language and style required
( ) Moderate English changes required
( ) English language and style are fine/minor spell check required
(x) I don't feel qualified to judge about the English language and style

	Yes	Can be improved	Must be improved	Not applicable
Does the introduction provide sufficient background and include all relevant references?	( )	(x)	( )	( )
Is the research design appropriate?	(x)	( )	( )	( )
Are the methods adequately described?	( )	(x)	( )	( )
Are the results clearly presented?	(x)	( )	( )	( )
Are the conclusions supported by the results?	(x)	( )	( )	( )

Comments and Suggestions for Authors

Review of manuscript remotesensing-311343 "A double-sampling extensi
German National Forest Inventory for design-based small area estimation
forest district levels" by Andreas Hill et al..

The manuscript presents a very well structured study on an approach to r
data from large scale national forest inventories (NFIs) usable for decision
on a much smaller scale, namely forest (sub-) districts. As such, it can ha
significant socio-economic impact, as well as a strong influence on sustai
forest management.

Existing terrestrial sampling data from the third German NFI are combined
double-sampling framework, together with remote sensing data (an airbor
LiDAR derived canopy height model and satellite derived tree species
classifications map). Design-based small area regression estimators are f
used for obtaining precise predictions on a small scale. The resulting redu
sample by up to 43% compared to the traditional estimators is remarkable

The paper is well written and impeccably organized. I would like to show r
appreciation to the authors because they made an effort to make this wor
to read, objective and interesting. As a non-native speaker, I am not overl
qualified to judge English style and language. However, I noticed some ra
translations from German to English, and would therefore suggest to ask
speaker check the manuscript.

In my opinion, and from a forest scientist's perspective, the paper present
work, and offers a solid contribution to the field. However, I am somewhat
if Remote Sensing is the "right" journal for the manuscript. Maybe, a journ
focus on forest science or biometrics would be more appropriate. Regardi
wider audience of Remote Sensing compared to forestry journals, I strong
recommend to add some additional explanations of forestry specific conc
(especially angle count sampling) to make the manuscript understandable
readers from other disciplines.

Summarizing, I recommend to accept the manuscript pending minor revis

Good job on an interesting paper!

Please find attached a list of suggestions and comments on what I consider can further improve the manuscript.

##

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”

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

L36: “standwise forest inventories (SFI)”?

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)

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.

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.

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

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 & 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 & Saborowski, 2014)

L46-47: Double-sample procedures have not only be 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).

L99 “The German NFI (Bundeswaldinventur, BWI)“ I think, the abbreviation introduced in the next line becomes more meaningful when giving the German name here.

L100 “The third and most recent”

L101 “4x4km” I would prefer “4km ´ 4km”

L102 “2x2km” same as above (2km ´ 2km)

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.

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.

L196 Please also provide reference to the R software itself, and not only to the package.

L273: I think this is the only occasion you use italics for "*RLP*"

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.

L295 "RLP" was already introduced as an abbreviation

L287: Instead of “45 Forstämter (FA)” I would use “45 forest districts (Forstämter, FA)”

L288: Instead of “405 Forstreviere (FR)” I would use “45 forest sub-districts (Forstreviere, FA)”

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?

L295 "2x2km" (see above)

Figure 1: "communal forests" (see above)

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.

Table 3: I think one digit for the percentage values should be enough.

Table 6: Please indicate for what variable point estimates are. I suppose it is volume of growing stock per area unit [$m^3 ha\text{-}1$]? And again, check the number of digits.

L454-458 I understand your intention and the necessarity 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 somwhat 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.

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.

L655-656: "Northwest German Forest Research Institution" It should be "Institute" instead of "Institution"

General comments:

Overall, the manuscript is a relevant contribution to bringing the forest inventory back to where it belongs, namely within the infinite population approach to sampling. Although the scientific contribution is not entirely novel - because the estimators have been described in detail in Mandallaz (2012) *"Design-based properties of small-area estimators in forest inventory with two phase sampling. Available at http://e-collection.library.ethz.ch.)"* and *Mandallaz (2015) "Mathematical details of two-phase/two-stage and three-phase/two-stage regression estimators in forest inventories: design-based Monte Carlo approach" (Available at https://www.research -collection.ethz.ch)"*, the authors provided a solid, "down-to-earth" piece of work, relatively easy to follow also by the non-sampling experts and practitioners. Nevertheless, there are a few issues that, in my opinion, deserve attention, and perhaps some clarifications should be added.

Comment 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 & 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.

Comment 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 & Bennett (1998) "Robust models in probability sampling. Journal of the Royal Statistical Society, 60, 3-21*".

Comment 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 SA 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 & Särndal 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.

Minor comments

Comment 1.
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.

Comment 2.
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 & Bennett (1998) "Robust models in probability sampling. Journal of the Royal Statistical Society, 60, 3-21"*.

Specific comments

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

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

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

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.

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.