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.