Dear Mr. Hill,  
  
Reviewers have now commented on your paper. You will see that they are advising that you revise your manuscript. If you are prepared to undertake the work required, I would be pleased to reconsider my decision.  
  
For your guidance, reviewers' comments are appended below.  
  
If you decide to revise the work, please submit a list of changes or a rebuttal against each point which is being raised when you submit the revised manuscript.  
  
Your revision is due by 05 Feb 2018.  
  
Please make sure to submit your editable source files (i. e. Word, TeX).  
  
To submit a revision, go to http://efor.edmgr.com/ and log in as an Author. You will see a menu item called "Submissions Needing Revision". You will find your submission record there.  
  
Yours sincerely  
  
Hans Pretzsch  
Editor-in-Chief  
European Journal of Forest Research  
  
COMMENTS FOR AUTHOR:  
  
**Reviewer #1:** This relevant and interesting study describes the development of a working/linking model to be used with model-supported estimators. The challenge of the study was the large spatial extent in combination with high resolution auxiliary variables and field data that resulted in severe inconsistencies that had to be handled. The applied approach is statistically rigorous and adequately described; the text is well written and clear.  
  
Comments  
  
Even though the development of the working/linking model is important, I would have appreciated if it was also presented what the effect of using the model is for estimates. How big is the relative efficiency for estimates in RLP? Consider discussing the impacts of using an internal model (as here) vs. an external model.  
  
Please discuss the results with the study by Kirchhoefer, et al. (2017) {Considerations towards a Novel Approach for Integrating Angle-Count Sampling Data in Remote Sensing Based Forest Inventories, Forests}  
  
Despite of the large number of observations, the number of 39 explanatory variables is quite large for a linear model. Could you discuss implications of this? Except for 2007, the model parameters do not seem to be all too different (Fig 7). Consider merging several years (all except for 2007?) into one factor as a simple means to reduce the number of model parameters.  
  
The use of square supports is at least uncommon. Basically all studies I am aware of use circular supports (in the case of circular sample plots as here). Therefore, this choice should be justified a bit more or be revised. The reason given that the support should allow for a potential tessellation does not seem to hold because the exact plot location will not allow for an alignment with a potential tessellation grid anyways. To my understanding, the area (size) of the support needs to fit to the field data and the tessellation grid, especially if scale-dependent explanatory variables are used. The shape of the support should resemble the field data. I think the model variance is artificially increased by selecting a support that does not fit (i.e. is not circular) with the field data.  
  
Was it not necessary to remove outliers or other influential observations from the data set? It sounds almost too good to be true, if that was not necessary.  
  
It is also somewhat uncommon to derive explanatory variables for CHMs instead from ALS raw data. A lot of information seems to be lost that way. Please justify or revise. Differing pulse densities usually do not have much influence on working models and can easily be considered in the model. This has probably a technical reason?  
  
Why and how were the ALS raw data thinned before interpolating them to grids? (Consider giving point densities in the more common unit point per m2.)  
  
Edge correction. Figure 3b suggests that supports were clipped at forest boundaries. 1) There is an additional data set for forest extent of public forests. Could it be described a bit more? Is the model, strictly speaking, only valid for public forests? How were plots on private forests treated? 2) Does clipping of supports fit to the type of edge correction in angle count sampling used in BWI3?  
  
Calibration. Did this study really "introduce a calibration technology"? Consider revising. Why is the tree-species model a "calibration model"? Calibration model sounds like the parameters of the original model were adjusted. Calibration is also an estimation technique and could be misunderstood in this context. Would it be a calibration model if a traditional tree species map was used as explanatory variables? Is it not simply a model with some categorical explanatory variables?  
  
The issue of using y-transformed models or not is of high importance. By discussing why g-weight variance is of important, this paper could contribute considerably to the discussion. In addition: How are negative predictions dealt with in practice? Set to 0?  
  
P1,L44 RHS: Is Beaudoin et al really the right reference here? The concept is anyways much older and Næsset, E. (1997) {Estimating Timber Volume of Forest Stands Using Airborne Laser Scanner Data. Remote Sensing of Environment} should be considered.  
  
P2,L15: Does Van Aardt et al. fit into the list of references?  
  
P2,L59: It may be of interest around that line that a variable describing deciduous proportion derived from leaf-off ALS data was used by Breidenbach et al. (2008) {Mixed-effects models for estimating stand volume by means of small footprint airborne laser scanner data. Photogrammetric Journal of Finland} to improve a model for timber volume.  
  
P2,L5-8 RHS: "One of the rare examples…" I cannot follow here. The way it is described, it is exactly the approach commonly used in the Nordic countries. There must be hundreds of studies one could cite. Probably a misunderstanding?  
  
P13,L31-37: Why blending ITC into this article? Does the concept of supports fit there at all? Consider removing.  
  
Consider discussing the model a bit more with other models published in central Europe (maybe based on smaller study sites) and other studies based on NFI data. Why do we see differences?  
  
P7,L30-31: Ambiguous. Please rephrase. Is dominant species here decided based on field data or the map?  
  
P7,L7-32: The model interpretation is a bit lengthy and may be best described by a reference to Fig 7.  
  
Pleas add a table on field measured values and relative RMSE values (%).  
  
Please consider that the Conclusion is not an extension of the Discussion. Consider revising to shorten and to avoid references in the Conclusion.  
  
Fig 4: Please define "ind".  
  
Is Fig 5 really needed?  
  
Figure 6: Could you expand in the text why two distinct clusters are visible? This remained a bit unclear. The dashed lines are hardly visible.  
  
Figure 8: Consider showing only the simple model and the final model. The others seem to be of little relevance in this context and just make the graphic difficult to read.  
  
Figure 7: (The fig appears after Fig 8.) The single panels are very small and difficult to read. Consider just showing one panel to exemplify.  
  
Consider revising the use of percent (%) vs. percentage points. My impression is that the terminology could have been used in the wrong way. (And R2 values are proportions, not percentages.)  
  
Terminology (p3, l16): Consider using ALS throughout the paper as the correct acronym for airborne laser scanning. (Or lidar which is consistent with acronyms like radar or laser.)  
  
Please add page range when citing books.  
  
Consider capitalizing the terms Figure and Table in the text.  
  
Eq1: are Y(x) and e(x) defined? Consider presenting table 2 close to eq 1 as the submodel names were undefined until table 2 was presented.  
  
P5L37: Is the threshold defined?  
  
P8L17-21, RHS: Discussion?  
  
Format: Consider submitting a one-column, two line spacing manuscript, such that line numbers are available to all lines.  
  
**Reviewer#2:**

It is a very interesting and useful approach and paper.

Your research question is well defined, extensively investigated and explained.

Regarding your figures please be aware, that printing or digital presentation will be very small. Try to use clear distinguishable colors and markers. I recommend removing the grey background in all figures.

Some phrases are quiet long and contain several aspects. Try to shorten your sentences, by separating them according to the different aspects or contents.

Comments in detail:

P3 line 46 Column 2: Did you measure the absolute tree height of every sample tree, or was it a estimation according to the measurements of a sub-sample of tree?

P3 Line 48 Column 2: Do you have any information about the type and positional accuracy of the used GPS or GNSS-Techniques

P3 Line 57/58: Why did you restrict to stat and communal forests?

P4 Line 24-26: In general the point density is given in [points/m²]

P4 Line 47 -51: How did you calculate the canopy height model, by subtraction?

P4 L1 Column 2: How did you define the forest borders?

P5 Line 34 – 40: Mention here, that you investigate this threshold in your work.

P5 Line 11-36 Column 2: Why did you use these variables and not others? I would recommend putting the explanation of your calibration method more popular in your article. It is an important part of your work.

P6 Line 46: the title should be “model building and evaluation” instead of model validation

P7 Line 4-18: What are the values for lidar year,   0 and 1 or N/A and 1 or anything else?

P7 figure4: the figure is overloaded. Try to use colored lines instead of colors and different markers. Is it really necessary to show both n and threshold? If yes, put them into two separate figures. Or could you explain some effects by n within your discussion. Anyway try to put the legend below or above your figure.

P8 Line 1: Title should be “Calibration” and not only “Effect of Calibration”.

P8 L55-59: Separate it in at least two sentences.

P 9 figure5: the figure is hard to differentiate:  Use color instead of different markers, put the legend below or above could already help.

P9 figure 6: Where do the different grey values refer to?

P10 figure 8:  I recommend limiting the y-axis between 0.2 and 0.6. A legend is missing and should be added. The long title of the figure contains text, which should be included in the text and not in the title.

P10 Table 1: I recommend mentioning RMSE % instead of SSE

P12 Table2: submodel 3 isn’t mentioned and explained in the text. It isn’t possible to reconstruct how you get your amount of parameters for each model. In addition the amount of parameters isn’t discussed in the text. So I recommend either to explain (I would prefer) or to leave it out.

P12 Line35: This is the wrong reference: Table1 instead of Table2.

P13 Line 43 Column 2: aerial instead of areal.