Author’s responses to revievers comments, 2nd round

essd-2022-152

Referee report 1 and authors responses

Referee 1 wrote:

The authors have adequately addressed my previous comments and I have just a few additional suggestions to address before the paper is published. A couple of these have to do with the data organization, and while I see the the pangaear package is convenient for accessing this data, I generally think there could be some improvements still. Specifically, I would recommend including a comment containing a brief description as well as the units for each variable in the parameter list. As it stands, some metadata information is included in the comments for some variables, while in other cases it is included in column headers. This inconsistency can make it hard to find things, and when column headers are used as variable names including acronym definitions and units makes them quite unwieldy. Sorry I didn’t catch these things in my previous review.

Our response:

Much of the variable naming in the data tables is due to PANGAEAs policy regarding the organization of data, which aims at using variable names that are more . To help the user to get an overview over the different variables, their meaning and units, we replaced the tables in Appendix B (which until then displayed all variables grouped by whether they are measured or derived). The new table includes the variable names from the *.tab files, the units, acronym definitions, and whether the variables were measured or derived, and if so, from which variable it was derived.

Referee 1 wrote:

The only thing that is more than editorial in nature is the fact that the data set contains repeated observations (mentioned on line 99). It seems that because of this the same tree has multiple unique tree ids - is this correct? I don’t think this is tremendously problematic because the plot-level data exists as a separate data set. On the other hand, I could envision a case for example, where a user wants to work directly with the tree level data to calculate other plot-level metrics and aggregates by plot (i.e. the ‘Event’ variable) thereby including multiple trees. While I realize it us up to the user to read the paper and understand the data, I think that including repeated observations in the data set is not good practice and would recommend either having a more explicit way to indicate which observations are repeated, or publishing the intensive circular plot tree measurements as a separate data set. And on a related note, for the trees included twice, do both records include measured DBH, or is one measured (i.e. CIRCLEPLOT) and the other predicted from height (e.g. PLOTHEIGHT)?

Our response:

Yes, it is true that the same tree can have two entries into the database with two different tree IDs. The two entries can generally not be matched to each other, because the height-only inventory and the detailed inventory were usually performed as two separate steps. Therefore, it is possible that the same tree appears once with measured DBH and once with DBH predicted from height via the allomeries.

In order to mitigate this problem, we have split the database into two tables, which we called “Tree heights”, and “Tree Measurements”, as the first one contains all entries with the survey protocol “PLOTHEIGHT”, and the other one all entries with diameter measurements. The entries with the protocol “PLOT” were copied to be present in both tables. This way, each table for itself does not contain any duplicate trees, “Tree Heights” contains all trees used for the aggregation of variables on the plot level, and “Tree Measurements” contains all entries needed for calculating the allometries.

Referee 1 wrote:

Also related to the data, it is unclear what the difference is between Longitude and Long C as well as Latitude and Lat C are. Appendix B indicates that Lat C and Long C are derived variables, but the methods or the metadata don’t seem to indicate how they were derived or how they differ from plot coordinates acquired in the field.

Our response:
The fields “Lat C” and Lon C” in the Tree Data Base indicate the coordinates of the plot center, and represent the exact same measurements as the plot coordinates in the Plot Data Base. The reason why they were considered to be “derived variables” in the Tree Data Base is that they were not measured individually for every tree, but copied from the Plot Data Base, and thus derived from the original “measured” variable “Event” in the Tree Data Base. In the new table in the appendix, which replaces the former one that categorized the variables as measured or derived, this is explained as

derived (from Plot Data Base)

Referee 1 wrote:
L151: should be ‘updated’, I think.

Our response:
Corrected accordingly.

Referee 1 wrote:
L199: I think this sentence refers to basal diameter, but it is a little unclear. Perhaps replace ‘It’ with ‘DBH’.

Our response:
‘It’ refers to diameter at breast height, as the range of basal diameter was already stated earlier in the paragraph. Adopted accordingly:

DBH is almost always lower than basal diameter, on average by the factor 0.628. DBH ranges up to …

Referee 1 wrote:
L235: Is this R2? It would be good to state this explicitly.

Our response:
Adopted accordingly:

… not exceeding an R^2 of 0.351 in any combination.

Referee report 2 and authors responses

Referee 2 wrote:
This is the second review of Forest structure and individual tree inventories of north-eastern Siberia along climactic gradients by Miesner et al.

The authors addressed almost all of my concerns and, with a few minor revisions, I think the manuscript will be suitable for publication.

General comments:

(1) The use of the terms ‘plot’ and ‘site’ interchangeably is unnecessarily confusing because those terms generally mean different things. Please chose one and use it consistently throughout.

Our response:
We had internally used the term “site” for describing the area of a plot and its immediate surroundings, but since this differentiation was not relevant in the manuscript, they became interchangeable, and we now replaced the term “site” by “plot” throughout the manuscript. The only exception is one occurrence, where there actually is a difference in meaning implied, which is hopefully understandable without further explaination:

The sites at which the surveys were performed were chosen beforehand with consideration of remote sensing data. […] The exact positioning of the survey plot was finalized on-site, with the aim to have each plot representing a homogeneous vegetation type.

Referee 2 wrote:
(2) The authors note in the main text that multiple regression is not a valid tool because the predictor variables are correlated among each other (line 251). I agree with this assessment, so I wonder why it is included at all in the manuscript. Either multiple regression analysis is appropriate (in which case, more needs to be done to ensure it is being used correctly), or it is not appropriate (in which case, it should not be included in the manuscript). The authors seem to want to have it both ways: they want to use the R 2 value from the multiple regression model while simultaneously calling such an analysis ‘not advisable.’ It is not statistically appropriate to use metrics from an analysis that has not been carried out correctly. I suggest removing all mentions of the multiple regression analysis from the manuscript.

Our response:
While the multiple regression may still give some insight about the data, even if its requirements are not fulfilled, we adhered to the suggestion and removed it. This is probably the most appropriate way of dealing with it.

Referee 2 wrote:
Line-by-line comments:

L64: It is not clear to me that an exception would necessarily be above, as the authors indicated in their response. If this is what you mean, change to ‘Annual precipitation is generally below 300 mm, although this is sometimes exceeded towards the boundaries of the area.”

Our response:
Adopted accordingly.

Referee 2 wrote:
L 77: For clarity, change to “the exact position of the survey plots was finalized on-site….”

Our response:
Adopted accordingly.

Referee 2 wrote:
L156: update should be ‘updated’

Our response:
Corrected accordingly.

Referee 2 wrote:
L223: ‘the’ is missing…should be ‘the significance and explanatory…’

Our response:
Sentence deleted (see below).

Referee 2 wrote:
L224: ‘value’ should be plural

Our response:
Sentence deleted (see below).

Referee 2 wrote:
L224: p-value of 0.021 is significant and R 2 =0.33 is fairly decent for ecological data, so consider removing the part about significance and explanatory values not being high.

Our response:
We re-formulated the paragraph to:

The Gini coefficients are negatively correlated with the geographic latitude of the plot (Figure 5). The linear regression has a p-Value of 0.021 and $R^2 = 0.33$.

Referee 2 wrote:
L311: In the previous review of this paper, I mentioned that one should be able to rearrange the equation relating DBH to height to compare data from the literature. The authors responded that this would be circular reasoning. Let me be more clear. The authors use the following equation:

DBH = a1*(H-1.3)^a2

We can rearrange this equation to relate height to DBH:

H = (DBH/a1)^(1/a2) -1.3

However, I briefly looked at the Alexander and Delcourt papers, and I was unable to find evidence that they relate DBH to tree height. Instead, they both relate DBH to biomass. I suggest the authors either compare the results directly or delete this sentence.

Our response:
It is true that the mentioned papers do not relate diameters to height, but to biomass. We therefore changed the sentence to:

There is little literature with which to compare our results, because commonly the diameter is used as predictor variable, and not height, as in Alexander et al. 2012 and Delcourt & Veraverbeke 2022, who both model biomass from diameter.

But we did not delete the sentence, because we find the two literature references worth citing in this context.
We did not do the rearranging of the equations as suggested, because inverting a DBH~H model generally does not lead to the same results as optimizing a H~DBH model, as the residuals are minimized along different variables.