General reply

Dear Referee,

We thank you for the critical assessment of our dataset paper. BorFIT is intended primarily as a training dataset for artificial-intelligence (AI) applications. These applications will thereby improve our understanding of the boreal forest's vegetation reorganization. BorFIT itself already enables 3D spatial analysis of species distribution and stand structure across the circumboreal region. The reviewed preprint provides detailed information on how BorFIT was created, validated, and what the strengths and weaknesses of the dataset are. We agree with the comments regarding the overall clarity and structure of the manuscript and will integrate them when editing an updated version.

Best

Jacob Schladebach

Our response regarding specific comments:

Data set and Pangaea

"Why the link to Pangea paper is included? I think only the reference is enough. Only the links to your dataset should be presented on the Data availability section to not create confusion. Actually, six links are presented which make a bit confusing how to actually find the BorFIT data set as a whole. Maybe the related five links could be referenced somewhere else and only on link to BorFIT data set should be clear presented on Data availability? Also would be nice to have some indication about the documentation related to the R code."

A readme will be added to the code repository. The "data availability" section in its current state includes the raw data used for BorFIT. However, since they are mentioned in Table 1, they can indeed be excluded from the Data availability section.

"When checking the Pangea repository and connecting here with the paper it is also good to include in the Data availability how the file is named and how the data was organized in the repository, for instance that they are presented in plot level. Also it is important to mention that the classes Trees and Species are extra bytes on Laz files and need to be considered when reading the files, since they are not standard fields on Laz structure. What is saving in the Classification and Point Source ID field? They look a bit strange. Please check."

The Pangaea repository includes a product guide which provides the requested information. Classification is the scalar field that indicates the ground classification values, based on LASTools (1 = unclassified, 2 = ground). Point Source ID corresponds to groups of points acquired on specific flight lines.

"Is EN23608(1)reference_plot_10_predicted.laz georeferencing correct? Looks that the trees are some how not vertical do the ground or the hill is very deep but the coordenates are somehow not aligned with E, N, up? Could the authors double check?"

The cloud is indeed located at a steep slope. I can not see any issues here.