

Dear Yuhan Rao,

We thank the reviewers for their thorough and valuable comments. We revised the paper according to the comments and recommendations of the two reviewers, see itemized replies to all comments and new text suggestions below.

reviewers comments in Arial
our responses in Arial bold

new text *“italic”*

Reviewer comment 1:

General Comments:

The dataset is relevant, especially due to its extended area and worth publishing. The data processing pipeline is standard and do not bring contributions. Regarding the ESSD data quality criterial, more clarifications are required from the authors. The Introduction and Methods are generally well described but would benefit from some clarifications. In the Introduction, certain statements could be revised to improve clarity and completeness. In the Dataset and Methods section, additional details on data acquisition, as well as on the final accuracy and spatial resolution of the dataset, would strengthen the manuscript. A major point to consider is adopting a clearer overall structure. It would be particularly useful for future users if the paper included dedicated sections on data validation and dataset usability. Finally, please double-check the Classification field in the point clouds, as some values appear unusual.

Thank you for your positive comments on the relevance of our BorFIT training data set and constructive feedback, which we have taken into account in our review. We clarified Introduction and methods, e.g. we edited the introduction to a clearer structure with more detail regarding common segmentation techniques and expanded the summary of existing data sets for similar purposes. We also added a validation section in “Methods”, “Results” and “Discussion” and a usability section in “Discussion”.

Specific comments:

Comment 1:

LiDAR is too broad terminology. I would recommend using UAV-LiDAR or UAV laser scanning system in the abstract to make it clear from the beginning which data collection method we are talking about here.

Response 1:

We thank the reviewer for the recommendation.

“UAV” was added to “LiDAR” in the Abstract and the Introduction to clarify.

Comment 2:

P1, Lines 27–30: I am not sure if I agree with the authors that this is the most used method nowadays. Some decades ago, yes, but nowadays there are many approaches, and stem detection has increased a lot, especially for boreal forests where the stems are often well visible. However, I agree that canopy-top or stem detection are the most common approaches for tree detection, although there are many other methods for segmenting trees depending on the spatial resolution, for instance cluster-based ones are also very popular. I suggest mentioning more than one methodology here (“the most used are”) and link with some review. There are reviews on tree detection and tree segmentation that could be used here.

Response 2:

We thank the reviewer for this recommendation. The introduction was restructured to account for the reviewer’s comment: we removed the statement on the ‘most used method’ clarified what segmentation techniques were popular, what weaknesses they have and what has been developed in the recent past.

“This requires the delineation of individual trees and their crowns with high spatial resolution remote sensing, also providing large area coverage. Passive, optical sensors are still widely used for this purpose. Approaches for tree segmentation on 2D true colour, infrared colour images or 3D derived canopy height models (CHM) from structure from motion (SfM) use area-based segmentation techniques to delineate the crowns and extract single trees (Polewski et al., 2015; Jaafar et al., 2018; Argamosa et al., 2016). Using these area-based methods, trees tend to be not optimally detected, especially when dealing with thin and overlapping crowns coupled with a dense understory layer, as we experienced for high latitude forests (Brieger et al., 2019). In contrast to passive optical imaging sensor acquisitions, light detection and ranging (LiDAR) is nowadays an active remote sensing method successfully used specifically in forestry.[...] With the optimisation of capturing 3D point clouds of forest structure from LiDAR and SfM,

specific techniques and algorithms for individual tree detection evolved. Point cloud voxelisation and deep learning are more recent approaches for individual tree segmentation from 3D point cloud data (Jaafar et al., 2018; Brunner and Houtmeyers, 2022; Zhang et al., 2023)."

Comment 3:

Line 33 - LiDAR can also be based on phase, not only on ToF, but I agree that ToF is the most common option. Maybe this should be clarified, since it is presented here as a theoretical explanation?

Response 3:

The Introduction mentions now briefly the full-waveform systems in an updated sentence, as follows:

"Discrete return LiDAR sensors, as used for this study, obtain detailed 3D point clouds of objects like trees and shrubs using laser pulses sent by the sensor. The pulses are reflected by a given object and the return-signal is recorded by the sensor. By measuring the Time-of-Flight (ToF), the distance to the object is calculated. This allows to differentiate between twigs, branches, vegetation and ground. Less common are LiDAR sensors that record the full waveform of the returning pulse. The data acquired by these sensors require more processing steps and are computationally demanding. However, they offer higher accuracy when detecting vegetation density below the canopy."

Comment 4-6:

Line 34 – "When properly segmented", "these point clouds". It seems that something is missing in this sentence, because point clouds were not mentioned in the previous sentences. Please revise the consistency and connection here. Also, the connection with the next sentence, which starts to mention tree detection, could be improved. Maybe you could include a sentence mentioning that to obtain individual tree point clouds automatically, tree detection and coarse-to-fine segmentation are often required. Should this be a new paragraph?

Line 35 -Well, biomass estimation with UAV is somehow possible, but there is still a lot of discussion about its accuracy. The UAV point cloud should be very dense. Even leaf and wood separation in UAV data is still quite challenging. I would reconsider this affirmation. Please check the results reported by Brede et al., 2019

Brede, B., Calders, K., Lau, A., Raunonen, P., Bartholomeus, H. M., Herold, M., & Kooistra, L. (2019). Non-destructive tree volume estimation through quantitative structure modelling: Comparing UAV laser scanning with terrestrial LIDAR. *Remote Sensing of Environment*, 233, 111355.

Line 38 – The author mentioned that “AI highly depend on the selection of appropriate algorithms,” which I agree with, but it also often depends on data training, which is most problematic and which is the reason for the paper. I think this should be included here too. This comes later on line 41 in the species classification topic, but it is also very important for tree segmentation. Also, the connection with the next process step could be improved (lines 39–40). I agree that segmentation has a huge impact on species classification, in which noise from neighboring trees can certainly intervene in the result. This was the main problem in the paper by Shcherbacheva et al., 2024. Could the authors elaborate more on this connection in the text before moving to the species classification topic?

Shcherbacheva, A., Campos, M. B., Wang, Y., Liang, X., Kukko, A., Hyyppä, J., ... & Puttonen, E. (2024). A study of annual tree-wise LiDAR intensity patterns of boreal species observed using a hyper-temporal laser scanning time series. *Remote Sensing of Environment*, 305, 114083.

Response 4-6:

The paragraph was adapted to a clearer structure, leading the reader from the recorded point clouds, via individual tree segmentation towards object-based analysis, e.g. species assignment:

“High quality tree segmentation is a prerequisite for object-based analysis like species classification. Segmentation quality is critical, as misclassifications in object-based analysis often stem directly from segmentation errors (Shcherbacheva et al.,2024). The performance of deep learning models for tree segmentation is highly dependent on selecting suitable algorithms and locally acquired training data, with dense and biodiverse forests posing additional difficulties (Falkowski et al., 2008; Hastings et al., 2020; Ma et al., 2022). Therefore, a high-quality training dataset with individual trees and species specifically tailored to the region of interest is essential to aid individual tree analysis.”

Comment 7:

In the dataset review, it would be interesting to also mention the FOR-instance dataset, which is a UAV laser scanning dataset aimed at semantic and instance segmentation of individual trees. Species information is also provided there. Overall, the review of available open datasets could be expanded. There are other datasets that provide individual trees, species, and leaf–wood separation, which could also be mentioned, even if they contain fewer samples. I think that extend the review here would improve the contribution discussion of this dataset and make more clear the differences between the proposed open dataset and existing ones, particularly in terms of coverage, spatial resolution, and temporal resolution too.

Puliti, S., Pearce, G., Surový, P., Wallace, L., Hollaus, M., Wielgosz, M., & Astrup, R. (2023). For-instance: a uav laser scanning benchmark dataset for semantic and instance segmentation of individual trees. *arXiv preprint arXiv:2309.01279*.

Response 7:

We informed on the FOR-Instance data set in the discussion. In this updated version, we also added information and reference in the introduction. Further, we expanded the comparison to other data sets also covering other world regions, as the availability of such data sets for the boreal region is limited. The new paragraph reads as follows:

“There are several forest structure data sets available for central European mixed forests to serve as benchmarking data sets for AI algorithms (Weiser et al., 2022; Henrich et al., 2024; Andreas et al., 2025). Weiser et al. (2022) published a data set with individual tree point clouds from 12 different forest plots located in Germany, combining UAV-LiDAR, terrestrial laser scanning (TLS) and airborne laser scanning (ALS). The data can be used to extract individual tree metrics or for radiative transfer modeling. TreeLearn (Henrich et al., 2024) provides both a data set of individual tree point clouds and an algorithm, that was trained on that data, specifically aimed at individual tree segmentation. Overall, there is good representation of such data sets for European forests with a considerably lower number available for other world regions. Chavana-Bryant et al. (2025) provide ForestScan a multiscale dataset from ground based tree census to UAV and airborne LiDAR point clouds with 3D forest structure of global tropical forests without individually segmented trees. ForestScan is not considered a training data set but provided as an validation data set for earth observation (EO) missions, rather than segmentation algorithm training. High-resolution data sets covering a diverse range of tree species and environmental conditions from the boreal region remain scarce. van Geffen et al. (2022) provided SfM point clouds of forest plots, which include CHM and orthoimages for the Eastern Siberian boreal forest, but not individual trees. The LiPheStream data set by Wittke et al. (2024) includes point clouds of 458 individual boreal trees from Finland and offers high temporal resolution through an 18-month monitoring phase, covering the phenological dynamics over two growing seasons. Specifically catered towards training AI applications is the FOR-Instance UAV laser scanning benchmark data set for semantic and instance segmentation of individual trees provided by Puliti et al. (2023). FOR-Instance contains individual tree point clouds from various regions, including the European boreal zone.”

Comment 8:

Also in the introduction (same comment as in the abstract) the author mentioned that BorFIT comprises “384 LiDAR point clouds collected from 146 locations across bio-climatic gradients in Siberia, Canada, and Alaska, with each point cloud containing up to 200 individual trees with assigned species”. It is very important to make clear here that the source of the data collection is a UAV laser scanning (ULS). This gives to the user the insight of what type of spatial resolution we are talking about in this dataset. Consider including this information also in the abstract. Consider also this comment on the Abstract presented at Pangea on the data server.

Response 8:

We included the information in the abstract and Introduction (also see Response 1).

Comment 9:

Line 53 - Typo error: tree species "classificaiton"

Response 9:

Thanks for pointing this out, it has been corrected.

Study region

Comment 10:

Line 61- The first sentence of the dataset description is a bit confusing "UAV LiDAR surveys in addition to extensive multisensor drone overflights". What this means? What type of multisensory? Laser scanning was not included on those platforms.

Response 10:

The sentence was misleading and was corrected as follows:

"Since 2021, we carried out UAV-LiDAR surveys in addition to drone overflights using other sensors, e.g. multispectral and hyperspectral, in various locations within the [...]"

Comment 11:

On the research area description and also on Table 1 would be important to add the exact extension of the area covered, "several km" is too vague. Also some coordinates limiting the area would be welcome to make the extension of the research area more clear.

Response 11:

The sentence was adapted. The length of the transects varies greatly, so the phrase 'several kilometres' has been removed. The coordinates marking the extent of each transect were added to Table 1.

Comment 12:

Are the reference plots managed forest? Looks like from the point clouds that some areas have very sparse vegetation. Is it natural?

Response 12:

All plots are natural forests. To clarify this, a sentence was added to the “Study region” chapter.

“All sites are natural, unmanaged forests.”

Comment 13:

What is meant here by “inaccurate trajectory strips”? Could the authors explain more clearly which criteria were used to consider a trajectory inaccurate? Was any checkpoint or other feature used to verify the accuracy of the georeferencing? What is the estimated georeferencing accuracy?’

Response 13:

“Inaccurate” was misleading in this case. It was changed to “unwanted”. The process of point cloud generation involved the selection of trajectories which recorded the data that we wanted to use. Trajectories of the initiation pattern were removed for example.

Due to often dense canopies and challenging accessibility of the terrain in the field it was not practical to use ground control points for the acquisitions. We relied on the GNSS data continuously logged during the flight. With post processing GNSS data with POSPac RTX, typical accuracies are approximately below 10 cm, compared to global geoposition.

Comment 14:

Line 77 – The first sentence of the data processing section is a bit difficult to read in the current format. Could the authors revise to make it clearer?

Response 14:

We adjusted the part for clarification and it reads now:

“UAVs (DJI Matrice M300 RTK) were used to conduct transect overflights covering at least 500 × 50 m. For these flights, a LiDAR sensor (YellowScan Mapper; Table A1) was employed. In addition, an RGB camera module was used during the 2022 and 2023 acquisitions.”

Comment 15:

Line 79-80 [The LiDAR mapper followed....] Could the author explain better this sentence? Is this information really needed? Instead other technical specifications that are presented as supplementary could be in the main text. I think the specifications of the GNSS/INS system used and the expected accuracy of the position and attitude estimation for the direct georeferencing is much more important here and should be clarified. Also how the laser scanning system was setup is not clear, which was the frequency used, angular resolution, FoV, angle of aperture, laser divergency, etc? The table with all the specifications of the sensors used are important for facilitate future users information.

Response 15:

We acknowledge that clarification was needed here. We decided to edit the sentence instead of removing it because it plays a significant role in data acquisition accuracy:

“The LiDAR mapper followed a predetermined initiation pattern for Inertial Measurement Unit (IMU) initialisation; this pattern is designed to ensure accurate calibration of the IMU sensors by stabilizing the device and minimizing drift errors before data acquisition begins.”

Comment 16:

Figure 3 is hard to interpret. Could the author colorize the point cloud by height for instance, since the main important feature is the size of the trees?

Response 16:

We agree that the figure was quite complex and thank the reviewer for good suggestions. We changed Fig 3 accordingly by adding the colour-coded point cloud of the flight transect, visualising the tree height, similarly, we colour-coded the reference plots. We also added the information in the figure caption that Fig 3 represents a schematic figure and added more explanatory titles in the figure itself. The figure was adapted to a more schematic format.

Comment 17-18:

Line 94 – the terminology reference plot “cut-outs” is a bit strange. I think it is enough to tell data 20x20 plots are selected or generated from the transects or maybe used the word segmented or sampled. Try to standardize the terminology.

Tree height groups 1 to 10 need to be better explained with quantitative values, small, medium, high do not say much without know the forest context. I suggest adding a table here with the range of heights in which level and which is the dominant species/ plot area in which of those. This can assist the users later to search better the data that would be useful for different purposes.

Response 17-18:

The term “cut-outs” was replaced by “segmented” or “segment”. Furthermore, following the recommendation of the reviewer, on PANGAEA, the metadata of the reference plot files were updated to include dominant tree species and average tree height. Figure 3 (see response 16) has been updated to include low, medium and tall trees as additional explanation to the reference plot numbers.

Tree segmentation

Comment 19:

The authors mentioned that The training data sets are based on the reference plots annotated with “00” where species were documented during a field survey, but it was not clear if this plot has an example of all species to be classified. If this is not the case, how this would affect the classification results if the samples are not significant? Later the authors also mentioned that they used distinct training data subsets and four random forest models, which it is ok but should be better presented since the begging. Now the sequence that the information is presented is a bit confusing and mix. I suggest to reformulate the explanation logic here.

Response 19:

We used two training data sets, one for North America another for Yakutia. Included are all reference plots “00” for the regions. Therefore, each training data set contains individuals of every species present in our North American data or Yakutian data. Since the location of the reference plots was chosen to be representative for the local forest conditions, it can be assumed that there are individuals of every species located in the proximity. However, it can happen, that the random forest classifier predicts species on the reference plot, that are not present there. Therefore, we included the likelihood of the prediction into the extra byte “prediction” for every tree.

To clearly state the use of two distinct training data sets, we edited the opening sentence of the paragraph:

“Two training data sets, one for North America, another for Yakutia, were created based on the reference plots annotated with “00” where species were documented during the expeditions.”

Comment 20:

It would be possible to assign and discuss some type of data quality regarding the data acquisition and segmentation? Like if the tree is full visible (stem and canopy boundary seems well represented) or if there are occlusions or if there is noise from other trees even with the manual segmentation?

Response 20:

We did not record this information. However, we excluded reference plots with very low point densities to ensure consistent data quality. As discussed in the manuscript we expect more occlusions on plots with broadleaf trees as overhanging branches aggravated manual segmentation.

Comment 21:

Figure 4 - which species is presented in Figure 4? This is very minor, but normally it is easy to present the individual trees with its height and not with absolute terrain values from the georeferencing just to have a more fast read of the tree size.

Response 21:

It is an individual of *Picea mariana*. The figure is an example plot of an individual tree point cloud. It is not meant to be interpreted in terms of a measure of quality control to identify outliers that might be included from the manual segmentation. Therefore, the point clouds have not been terrain normalized. This is now stated in the “Data acquisition and postprocessing” section:

“The point clouds have not been terrain normalized to keep the information in their true-dimensional space with height values on the Z axis representing above sea level”

Comment 22:

How this eleven structural and two spectral variables was select? Could the authors link with previous works that also performed Species classification to give more support to the method and features choose to be used here? This geometric variable seems a bit unusual selection, since most part of it focus on height and volume of the canopy. Maybe would also be nice to discuss the contribution of this features for the classification? You did not include LiDAR intensity? Normally it is a very good information to separate desiduos and coniferous trees.

Response 22:

The variables that we used for classification are the result of iterative testing. During this process, we tested different sets of variables, always checking the model output and the individual variables contribution. Intensity, in our experience, did not significantly contribute to the model's prediction accuracy. The geometric, volume based and point distribution-based variables are inspired by studies from Vauhkonen et al. (2009), Irwin et al. (2024) and Qian et al. (2023) as mentioned in the manuscript. The spectral variables are based on what's possible with only RGB data. The normalized green red difference index (NGRDI), has been tested as a proxy for NDVI by Al Rosyid et al. (2022). The variable atmospherically resistant index (VARI) has been proven to be sensitive to the

discrimination of plant growth stages and vegetation health. The citations in the manuscript were updated.

Andrade, R. G. et al.: Monitoring of corn growth stages by UAV platform sensors, *International Journal of Advanced Engineering Research and Science*, 6, 54–58, <https://doi.org/10.22161/ijaers.69.5>, 2019

Al Rosyid, A. N. I., Astika, I. W., Setiawan, Y., Muttaqin, K. H., Impron, Imantho, H., Sugiarto, S. W., Endiviana, O. A., and Yuliawan, T.: Visible Band Optimization of Unmanned Aerial Vehicle for Estimating Synthetic NDVI on Rice Vegetation, *Jurnal Keteknikaan Pertanian*, 10, 281–290, <https://doi.org/10.19028/jtep.010.3.281-290>, 2022

Vauhkonen, J., Tokola, T., Packalén, P., and Maltamo, M.: Identification of Scandinavian Commercial Species of Individual Trees from Airborne Laser Scanning Data Using Alpha Shape Metrics, *Forest Science*, 55, 37–47, <https://doi.org/10.1093/forestscience/55.1.37>, 2009.

Qian, C., Yao, C., Ma, H., Xu, J., and Wang, J.: Tree Species Classification Using Airborne LiDAR Data Based on Individual Tree Segmentation and Shape Fitting, *Remote Sensing*, 15, 406, <https://doi.org/10.3390/rs15020406>, 2023

Irwin, L., Coops, N., Riofrío, J., Grubinger, S., Barbeito, I., Achim, A., and Roeser, D.: Prioritizing commercial thinning: quantification of growth and competition with high-density drone laser scanning, *Forestry: An International Journal of Forest Research*, <https://doi.org/10.1093/forestry/cpae030>, 2024

Comment 23:

I'm not sure that the traditional section titles, such as Results and Discussion, are the most appropriate for a dataset paper. It is important to keep in mind that this paper will serve as a reference for future users of the dataset, who will rely on it both to understand the data processing steps and to cite it in their own work. Therefore, I suggest reorganizing the structure to better align with the goals of a dataset description paper. Many dataset papers diverge from the traditional Results/Discussion format. Adopting a similar approach would benefit this paper as well. For instance as: <https://doi.org/10.5194/essd-17-4569-2025> and <https://doi.org/10.1038/s41597-024-04143-w>

A possible outline could be:

1. Introduction
2. Methods and Data Processing – detailing all steps used to produce the final dataset.
3. Data Validation – evaluating the dataset's accuracy, including classification performance and the related conclusions currently placed in the Discussion. I think the Class errors presented as supplementary are super important and could be in the main text.

4. Dataset Properties – presenting the content from Sections 3.1 and 3.3, which contain excellent visualizations. These could be reframed as a description of the dataset from an ecological/forest perspective.
5. Usability – a section that is currently missing, which would explicitly discuss and clarify the potential applications of the dataset, based on both the technical properties (from the validation) and the ecological/forest context (from the dataset description).

Regarding data validation, the species classification results could be discussed in more detail, clearly stating which classes users should be more cautious with when using the dataset. Including a table with estimated performance per species could be very helpful. From the data processing point of view manual segmentation is often the most accurate approach and I do not have much to comment about that, but it would also be valuable to provide additional details on how spatial resolution varies and data acquisition influence the dataset.

Response 23:

Thank you for your recommendations. We would like to keep the ESSD standards with Introduction, Data and methods, Results, and Discussion. We highly appreciated your point that validation is a key information of a derived product such as the BorFIT data set. We also agree that our validation methods and results did not come out clear in the first version of our manuscript, Therefore, we substantially included more information on the validation, also resulting in an update of the structure and chapter names. A new “Validation” subsection is now part of 2.Methods, 3.Results and 4.Discussion. We also added a “Feature Importance” subsection in 3.Results and a “Usability” subsection in 4 Discussion where we describe the potential applications of the data set..

Comment 24:

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.

Response 24:

We agree with the reviewer that the data availability section was not yet clear. We enhanced the overview and accessibility of the BorFIT data set and its source data by displaying the links and references in a table for a better overview.

Comment 25:

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.

Response 25:

We thank the reviewer for pointing this out. Within the “Data availability” section, we now include a description of the file names and we now point out the necessity to load the added scalar fields as extra bytes. A product guide, that includes this information is added to the PANGAEA repository. The field classification is the result of the ground classification that was performed using LASTools, as mentioned in the manuscript (1 = unclassified, 2 = ground). Point Source ID corresponds to groups of points acquired on specific flight lines.

The paragraph now reads:

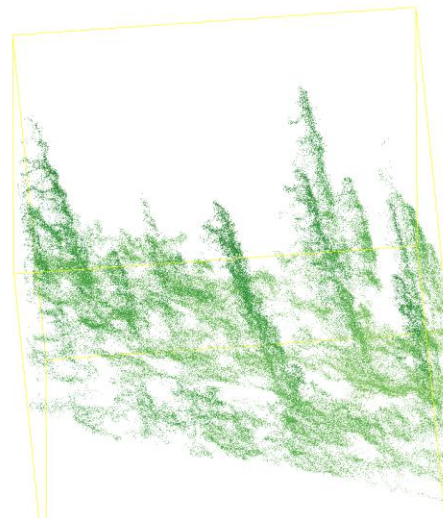
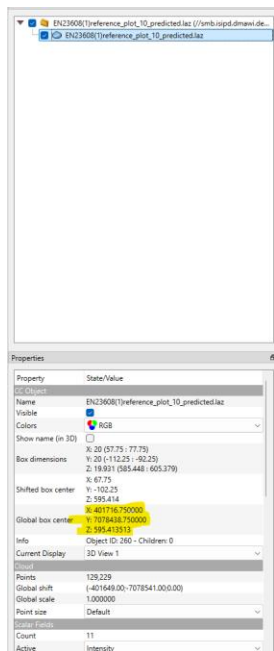
“The BorFIT data set is published on the PANGAEA data repository (Felden et al., 2023). The R code used for training the random forest classifier and predicting the species is publicly available on GitHub (<https://github.com/StefanKruse/BorFIT/tree/main>). The files are named based on our expedition event number, representing the site of the drone overflights, e.g. EN23604_Lidar. "23" is the year of the expedition, in this case Alaska 2023, while 604 is the site number. The reference plot number, e.g. "00" is also indicated in the filename ("reference_plot_00_predicted.laz"). The added scalar fields "Species" and "Tree" are extra bytes in the .laz files, which needs to be considered when loading the data.”

Comment 26:

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?

Response 27:

The slope at the mentioned reference plot is indeed steep, since the point clouds are not terrain normalized, the trees are in this case not vertical to the ground. Coordinates are aligned correctly (see screenshot)



Comment 28:

The first authors of the dataset do not match the paper authors list. This is not a issue on my point of view but maybe the authors should consider to clear state the contributions of each author.

Response 28:

This is intended. Yes, we build up the author teams differently. In the author team of the data publication also people are included who contributed to the fieldwork, e.g. in terms of logistic support that are not part of the author team of the manuscript. In the credits we describe the contribution of each member of the author team of the manuscript.

Reviewer comment 2:

This paper introduces an interesting data set of segmented trees with species identifications in boreal forests across North America and eastern Siberia. While this is not a novel area of research and the paper makes no methodological innovations, having additional model outputs for this task may be beneficial to others looking to harmonize data from multiple acquisitions for their own modeling.

I have significant concerns (and confusion) about the combination of data with and without spectral information into a single data set. Given the two continents mapped here were modeled separately, the significantly different forest compositions across continents, and most North American sites had spectral information while the Siberian sites uniformly lacked it, it seems to me like users would be better served by having two distinct data sets made available. I don't think that this would need to significantly change the approach taken here, but rather treating "BorFIT" as the modeling approach taken and then having "BorFIT Siberia" as a distinct data product from "BorFIT North America" would possibly make the modeling and data processing steps used clearer.

I broadly agree with Reviewer 1's comments on alternative paper structures.

Thank you for your positive comments on the potential benefits of our data set. We understand the concern regarding the combination of data with and without spectral information within one data set. However, we decided to keep it combined, as the basis of data acquisition (flight parameters, LiDAR sensor) and processing are the same. The difference lies only in the added RGB camera module since 2022. To make differentiation and subsetting the data to the users needs easier, according to the reviewers comment, we updated the meta data table, to include the model used for prediction and the region for each file. That way, the users are able to easily identify the variables used for the prediction process.

Specific comments follow.

Comment 1:

Line 6: specify what type of lidar (UAV?)

Response 1:

To clarify this according to the reviewer's comment, UAV has been added in the abstract and the introduction.

Comment 2:

Line 11: Think this should be "random forest". I personally object to classifying RF models as "AI" but I understand that's the fashion these days.

Response 2:

It has been standardized to “random forest”. Thank you for your comment. We assigned Random forest as a machine learning model, to the subfield of artificial intelligence (AI).

Comment 3:

Line 25: Consider rephrasing this sentence. "Ecological attributions" is a particularly odd phrase.

Response 3:

As suggested by the reviewer, Attributions has been changed to contributions.

Comment 4:

Line 27: I disagree that this is the most common approach for segmentation. There's plenty of research on satellite, airborne, and UAV lidar for segmentation, and even imagery-based approaches are more likely to use hyperspectral imagery than just color and IR.

Response 4:

Thank you for your comment. The paragraph has been edited to include more recent approaches like point cloud voxelisation and deep learning. It now reads:

“[...] With the optimisation of capturing 3D point clouds of forest structure from LiDAR and SfM, specific techniques and algorithms for individual tree detection evolved. Point cloud voxelisation and deep learning are more recent approaches for individual tree segmentation from 3D point cloud data (Jaafar et al., 2018; Brunner and Houtmeyers, 2022; Zhang et al., 2023).”

Comment 5:

Line 30: What do you mean by "large areas"? I'd suggest UAV based lidar is probably the second least effective remote sensing approach (after only terrestrial lidar) for large area monitoring given range and speed limitations.

Response 5:

The sentence was removed in the updated version of the introduction.

Comment 6:

Line 40-48: I was confused by this segment. You appear to be attempting to describe why the boreal region presents special challenges, but none of these thoughts coalesce into a coherent explanation. Naming individual data sets in this context is similarly confusing. Consider rephrasing to make the challenges of this region more obvious. Suggest introducing BorFIT in a new paragraph.

Response 6:

The updated version of the introduction is largely restructured. It focuses on introducing different segmentation techniques including their weaknesses, then it introduces training data sets focusing on other world regions. This key message is, that such data sets are missing for the boreal regions of North America and Siberia. Areas of significance for global carbon balances and vegetation climate feedbacks. This is what we address with BorFIT, as now clearly stated in the introduction.

Comment 7:

Fig 1: Capitalize "georeferenced". Consider resizing elements to not break words. "randomForest" should be two words. Fig 2: Great map. Consider a box behind the legend for legibility.

Response 7:

As suggested by the reviewer, we edited the figures accordingly.

Comment 8-9:

Line 71: Please provide more detail about how sites were positioned. Were there formal criteria used or was this more ad-hoc?

Table 1: I have a lot of questions about how sites were selected. Why the large variation in number of point clouds and reference plots per region? Why was Alaska sampled twice?

Response 8-9:

ESRI satellite imagery was used to manually delineate transects representing different forest types and structural conditions. In North America, transects were established along highways and during multiple field expeditions spanning the southern mixed forests to the northern treeline and tundra. A Canadian transect was established in 2022, focusing on the tundra–treeline ecotone. Subsequently, an Alaskan transect was established in 2023, covering forest and mountain treelines across the Southcentral, Interior, and Far North regions. In 2024, fieldwork was extended to western Alaska, south of the Seward Peninsula, to investigate the westernmost forests and treeline transitions. The 2024 expedition also included revisits to previously established sites in Alaska's Interior and the establishment of additional plots in that region. Alaska was therefore sampled in two consecutive years; however, the BorFIT dataset includes observations

from only one time point per plot, and the 2024 data were independent of those collected earlier.

The following sentence was added to clarify:

“Alaska was therefore sampled in two consecutive years; however, the BorFIT dataset includes observations from only one time point per plot.”

Comment 10:

Line 95: Can you explain why you used this approach for selecting reference plot locations -- the stratification, 20m plot size, and number of plots selected? I don't have any concerns about the approach, but explaining the rationale (or citing to where this method was introduced, if appropriate) would be beneficial.

Response 10:

Our strategy was to group the existing information per overflight transect by tree height to ensure, that the different stand structures are depicted in our data. We extracted 10 reference plots from each point cloud to ensure good coverage of the different tree heights present. We started segmenting plots 1, 5 and 10 to cover both the smallest, tallest and medium sized trees present. The other plots 2-4 and 6-9 were meant for gap filling after 1,5 and 10 were segmented for each of our point clouds and were provided for selected transects when relevant.

Comment 11-13:

Line 110: Should be "issues arose"

Page 8 last paragraph: rather than "the same is true for *_Larix_*" consider "while *_Larix_* was the most abundant genus in Yakutia"

Line 135: Should be "random forest". Should cite Breiman 2001.

Response 11-13:

Has been edited as suggested.

Comment 14:

Table 4: Please provide citations for these indices and metrics.

Response 14: The table has been updated, now including citations for the variables if applicable.

Comment 15:

Line 153: Please clarify: was species composition predicted on structural parameters, or were the RGB data predicted?

Response 15:

To clarify the process, this sentence was updated as follows:

“The according tree species were instead predicted using a random forest classifier that was based only on structural parameters.”

Comment 16:

156: "random forest". If you're using the randomForest R package, you should cite it in this section.

Response 16:

We used the caret package and cite it in chapter 2.7.

Kuhn, M.: Building Predictive Models in R Using the caret Package, Journal of Statistical Software, 28, 1–26, 435 <https://doi.org/10.18637/jss.v028.i05>, 2008

Comment 17:

211: Did you manually segment 16,530 trees? You should mention this number earlier, in the "manual segmentation" methods section. If these are instead classified by the models, please define what "successful" means -- you only view about 4500 trees as being classified with a high enough probability to include in graphs in this paper.

Response 17:

Yes, we manually segmented 16,530 trees. This was now added to the “manual segmentation” chapter. The subset based on “probability” >0.8 was chosen because it reduces the noise of misclassifications. The analysis of species distribution based on higher quality predictions allowed to analyze whether the predicted distribution aligns with our experience during field work and reported habitats.

Comment 18 - 19:

Line 252: Remove "mean accuracy 82%", these models cannot be applied to the same geographic areas or even to the same subsets of data collected for this publication, and as such

their ensemble accuracy is irrelevant. "Accuracies ranging from 73% - 91%" seems more appropriate.

Line 283: "The same" (not Same)

Response 18:

We thank the reviewer for pointing this out, the manuscript was edited accordingly.