

Earth Syst. Sci. Data Discuss., author comment AC1
<https://doi.org/10.5194/essd-2021-281-AC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC1

Femke van Geffen et al.

Author comment on "SiDroForest: a comprehensive forest inventory of Siberian boreal forest investigations including drone-based point clouds, individually labeled trees, synthetically generated tree crowns, and Sentinel-2 labeled image patches" by Femke van Geffen et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-281-AC1>, 2022

Dear editor, Thank you very much for your time and effort. The review we received is included below in regular text and the changes that are included in the manuscript and the responses are written below each point in bold and cursive starting with the word response.

I, Femke van Geffen would like to, also on behalf of my team, thank the reviewers for their time and detailed comments on the manuscript and dataset. The comments greatly improve the work and are well appreciated. Kind regards, Femke van Geffen

- General comments:

These look like valuable datasets to make publicly available, from a vast region where there is a real paucity of data by any measure. These data applied to machine learning and other automated image analyses and interpretations should have value for remote sensing studies in the circumboreal region. I encourage publication, and could only note a few minor issues detailed below.

- Significance
 - Uniqueness

These data are definitely unique in that they are from Siberia, a larger expanse of tundra–taiga than in N. America, and yet almost all available ABOVE datasets are based in N. America, especially Alaska.

- Usefulness.

Is it useful or necessary to display images that were excluded from the dataset?

Response: This is included for informing on the quality of the data sets and to show details on our data processing standards: It shows the steps that were taken to make the dataset useful. Taking the images out that include people eliminates noise from the data such as the image shown in Fig. 8.

We also added the sentence:

L317-319: "The first fifteen images were excluded because they often contain a visual representation of the research team (for example, Fig. 8). Excluding these images reduces noise in the dataset as we aimed to include only forest and natural terrain in the images."

Figure 1. The study region names Chukotka and Yakutia are missing from the map or caption.

Response: Thank you for pointing this out, the regions are now included in the figure caption below figure 1.

Table 1. If my math is correct, it looks like Lake Ilirney should have 27 plots not n=25 as stated.

Response: This is correct there are 27 plots. The mistake was made in the Bilibino plots where plot EN18033 and EN18035 are actually not included due to lack of usable data for these plots. That led to a mistake in the total which led to a mistake for the plots of Lake Ilirney, which in fact does have 27 plots. It is now corrected in Table 1.

- Data quality

L243. Were the cardinal directions adjusted for magnetic declination in the field, or prior to preparing the data product? This will affect how the plot sample area intersects with any georectified remotely sensed data.

Response: The magnetic declination in central Yakutia and Chukotka is very low (around or less than 10 °). We oriented the vegetation plot by iteratively moving with GPS with the set up for the true North, not the magnetic North, and did not use a compass for the North orientation of the vegetation plot. Like this, despite also the challenge of accurate GPS measurements in the field, we consider the North-orientation of the vegetation plot as robust and reliable.

Presentation quality

Fig. 6. The background RGB image is shifted between the images on the left and right, making it unnecessarily difficult to visually compare the overlaid features and labels.

Response: Yes thank you, The figure was redone and can be seen in the revised manuscript. Two figures were combined into one new Fig. 6, hopefully a clearer image.

- Specific comments:

L244-245. Could tree density per plot be calculated from these data? It sounds like the answer is no, if the only requirement was to sample 10 trees. How were the 10 trees selected from the size and spatial distribution of available trees in the plot?

Response: Yes, tree density per plot can be calculated from the data. The 10 trees were only selected to cover the whole size range of each species and investigate in-depth. All trees on the plots, with a height of >40 cm, were recorded and their height, vitality status, and species noted. This information per plot can be easily used to calculate the density.

In addition, the point clouds contain all trees on the plots. This information can also be extracted.

L556. Were the estimations made in the field or in the office based on visual interpretation of imagery?

Response: Thank you for your feedback. We have clarified it in the text. The recorded tree crowns refers to the trees measured in the field that were then linked manually to the orthoimages based on their locations and visual appearance. L410-415:

“In order to make assumptions and predictions about the content of the vegetation plots it is important to link the labelled individual trees from the fieldwork to the processed orthoimages. We located 872 trees and large shrubs in the orthoimages that were surveyed in Siberia during the two-month fieldwork expedition in 2018 (Kruse et al., 2019) (Fig. 16).”

The automatically detected tree crowns were estimated with detection software in R on the UAV orthoimages. Explained in the text at L273-279: “ The SiDroForest data collection also contains 19 342 automatically detected tree-crown polygons (Kruse et al. 2021b). The tree crowns were captured in the CHM by watershed segmentation analysis using the R package ForestTools (Plowright, 2019) and successive automatic generation of a polygon around them following Brieger et al. (2019). This automated tree-crown detection algorithm was run for all plots and the resulting shapefiles are provided for each plot that contained trees. Quality assurance was performed for each plot by carefully examining each plot based on expert knowledge and assigning a quality score of Q1 (good quality), Q2 (medium quality), or Q3 (poor quality) to the shapefile products.”

- Technical corrections:

L82. Not just a C sink, but also potentially a C source.

L93-94. This sentence is incomplete.

L103. The word “Likeness” actually isn’t and should be deleted.

L128. Change “was” to “were”.

L129. Change “was” to “were”.

L153. The sentence that ends here needs a period.

L203-204. State why datasets one and two were prepared before doing this for datasets three and four.

L240. Change “projective” to “projected”.

L253. Change “are” to “is”.

L618. Should not the citation be Fig. 22 (instead of Fig. 23)?

L672. Change “as” to “a”.

Response: All changes were made based on the technical corrections suggested.

