

Earth Syst. Sci. Data Discuss., referee comment RC3
<https://doi.org/10.5194/essd-2021-281-RC3>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2021-281**

Anonymous Referee #3

Referee 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-RC3>, 2022

General comments

The author collected tree - and plot-level forest structure data based on unmanned aerial vehicle and field investigation in two vegetation transition zones of Siberia, Russia. The datasets, including field plot level individual tree and shrub records (tree height, crown diameter, and species) and UAV products (e.g., Canopy Height) can be used for calibration and verification of model output, experiments, or observations. They are useful and important for future carbon dynamic studies and help to inform forest management, especially for the area where historical records and monitoring tend to be scarce. However, the manuscript is poorly organized and difficult to read. I encourage publication after addressing the following issues.

Significance:

The data are useful, complete, and fill in the region field data gap for Siberia boreal forests.

Data quality:

The data are easy to understand and presented readily and accessible to be used in other studies.

Presentation quality:

The manuscript was not logically articulated and was poorly written.

I recommend reframing the Introduction section. It should be more concise. I suggest shortening the detailed SiDroForest dataset and collection method description, but introducing the necessary or implication of each dataset (such as tree-level or plot-level records or canopy Height) and discussing the importance and challenges to collect these data.

The results seem like a duplicate of the method. I suggest including some further information and analysis, not only what the data were included or collected in the four datasets. For example, the frequency or distribution of tree species in the area for 3.2 Dataset 2, and the dominant vegetation classes and their distribution for Dataset 4 can be described.

Specific comments:

Line 68: change "are" to "is"

Lines 106-109: The sentence is difficult to read.

Lines 119-121: delete one of the "use" in the sentence

Line 124: add "which was" before "derived from"

Line 129: change "was" to "were"

Line 150: add ", " before "and"

Lines 243-244 and Figure A4: Two 30-m-long tape or 15m?

Lines 243-248: why were a minimum of ten individuals selected per plot? Why were plots recorded differently? If only part of the trees and shrubs were recorded, the data may be not able to represent the real forest information.

Line 252 please clarify the 11 vegetation classes here.

Lines 256-259 and Figure A4: The vegetation plot looks smaller than 30m ×30m (the double grid) if the red line is 15m long.

Line 274: with very low vegetation?

Line 295: add "that was" between "area" and "not"

Lines 301-304: Please break the long sentence into several simple sentences.

Line 326: add "that were" before "corrected"

Lines 338 -340: the sentence is difficult to read. Does that mean the tree crowns were captured by two methods: 1) watershed segmentation analysis and 2) successive automatic generation of a polygon around them?

Line 542: Link or Linking?

Lines 553-557: what's the meaning of this paragraph? Do you compare your field-measured crown diameters and detected crown polygons? If so, what's the difference between your results and the results of Bieger et al.?