

Earth Syst. Sci. Data Discuss., referee comment RC2  
<https://doi.org/10.5194/essd-2021-281-RC2>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## **Comment on essd-2021-281**

Anonymous Referee #2

---

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-RC2>, 2022

---

### Review Comments

The datasets presented seems to be quite useful for an important region where data is scarce for research. I believe these field survey data would be helpful for both small-scale and large-scale studies in understanding the circumboreal region. A few comments listed below for consideration.

Q1: Is the data set significant – unique, useful, and complete?

I think the data is unique and complete, though the authors can mention more about the usefulness of these datasets. For example, some of scientific questions that these datasets can help address? What specific applications can be expected from these datasets? Or how these datasets may help change the research landscape over time?

Q2: Is the data set itself of high quality?

Yes, the data is of high quality.

## Presentation quality

I do not think the authors pointed out specific/suitable software for simple visualization and analysis as I typically need to install some software to view the data.

## Specific comments

(1) The abstract can be improved by why these datasets are needed.

(2) The abstract seems to long and I would suggest further summarizing it and highlight the datasets in brief.

(3) I am not convinced how these field level data can help studies in this region. Only for optimized data containing annotated vegetation categories. What is the motivation for doing these machine learning applications? DO we need to look at data from a temporal perspective?

(4) The introduction should be shortened as well to highlight the contribution of datasets to the scientific literature or specific questions/challenges to be addressed

(5) Figure 1 suggested sampled sites are very limited. Can the authors justify why these sites were selected rather than other sites?

(6) I am not sure whether the codes for generating these datasets are available as it is not mentioned explicitly in the manuscript.