

Earth Syst. Sci. Data Discuss., referee comment RC1
<https://doi.org/10.5194/essd-2021-281-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2021-281**

Anonymous Referee #1

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-RC1>, 2022

- 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?

- Completeness.

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

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

- Data quality

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

- 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.

- 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?

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

- 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".

