

Earth Syst. Sci. Data Discuss., referee comment RC1
<https://doi.org/10.5194/essd-2022-296-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on **essd-2022-296**

Anonymous Referee #1

Referee comment on "A 29-year time series of annual 300m resolution plant-functional-type maps for climate models" by Kandice L. Harper et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2022-296-RC1>, 2022

Summary

This study describes a method used to produce the fractional composition of 14 PFTs at 300m resolution for the ESA-CCI land cover maps over the 1992–2020 period. Several 30m resolution datasets including surface water, tree canopy cover and height are used in the process, which is a significant effort given the large data volume dealt with at the global scale. The authors also compare the use of the new PFT data versus the previous version (based on a generic cross-walking table) in two land surface models for model simulation (ORCHIDEE) and evaluation (JULES), to demonstrate the impact of the new PFT data.

Overall the paper reads well, and the new PFT product is potentially very useful to the climate/land surface modelling community. However, I have a major concern about the accuracy and consistency of the high-res datasets used to derive the fractional composition of the PFTs, especially when the data values (e.g. tree cover and surface water products) are directly used to produce the PFT fractions at 300m pixels. Details are outlined below, which will hopefully improve the future version of the paper.

Major comments

(1) In the PFT product, the percentage of tree cover at the 300 m pixel is estimated using the 30 m tree cover data for 2010 from Hansen et al. (2013). Thus the accuracy of the tree cover in the PFTs is directly linked to the accuracy of the 30 m data from Hansen et al. (2013). Several previous studies showed that compared to field and other data sources (e.g. Lidar) tree cover data from Hansen et al. (2013) overestimated tree cover in their studied regions.

Tang et al (2019) showed that in the Sierra national forests USA the tree canopy cover from Hansen et al (2013) overestimated tree cover with RMSE around 20% when compared to field measurements, and with RMSE nearly 30% when compared to airborne Lidar estimates. Potapov et al (2015) found that the tree cover product from Hansen et al. (2013) overestimated tree canopy cover within the peat bog areas in Eastern Europe. They had to define "forest cover" using a tree canopy cover threshold of $\geq 49\%$. Wang et al (2019) also showed that tree cover from Hansen et al (2013) was overestimated in wetland environments over Canada. Though the tree cover data used in Potapov et al (2015) and Wang et al (2019) was for the year 2000, it was produced using the same method as for the 2010 data.

Tang, H., Song, X.-P., Zhao, F. A., Strahler, A. H., Schaaf, C. L., Goetz, S., Huang, C., Hansen, M. C., and Dubayah, R.: Definition and measurement of tree cover: A comparative analysis of field-, lidar- and landsat-based tree cover estimations in the Sierra national forests, USA, *Agr. Forest Meteorol.*, 268, 258–268, <https://doi.org/10.1016/j.agrformet.2019.01.024>, 2019.

Potapov, P., Turubanova, S., Tyukavina, A., Krylov, A., McCarty, J., Radeloff, V., and Hansen, M., 2015, Eastern Europe's forest cover dynamics from 1985 to 2012 quantified from the full Landsat archive: *Remote Sensing of Environment*, v. 159, p. 28-43, at <http://www.sciencedirect.com/science/article/pii/S0034425714004817>.

Wang, L., Bartlett, P., Pouliot, D., Chan, E., Lamarche, C., Wulder, M. A., Defourny, P., & Brady, M. (2019). Comparison and Assessment of Regional and Global Land Cover Datasets for Use in CLASS over Canada. *Remote Sensing*, 11(19), 2286.

Therefore, these uncertainties in the tree cover data would propagate to the derived PFT product. I wonder whether the Hansen et al (2013) tree cover data and the proposed method are best for producing the fractional composition of PFTs? Two other 30m tree cover datasets (NLCD and GLCF) were included in Tang et al (2019), though they did not perform better than the Hansen et al (2013) data in the evaluation. I wonder would an ensemble approach using all three of the 30m tree cover datasets as inputs for producing the PFTs be better? Just a thought.

(2) L576-578, "Since the PFT local product is built mainly for application to land surface models, the actual presence of grass vegetation vs. bare soil will be determined by the model given simulated or prescribed local climate conditions." This is likely the case when vegetation cover is dynamically simulated (with competition between PFTs) in the models. However, that is not always the case especially considering that prescribed PFTs are in general more realistic than dynamically simulated ones. For example, the majority of models participating in the TRENDY (trends in net land atmosphere carbon exchanges) project use prescribed PFTs without competition between PFTs in their simulations, which contribute to the annual Global Carbon Project's analysis of the land carbon sink (Friedlingstein et al., 2020). The simulations by ORCHIDEE in this paper also use prescribed PFTs. As demonstrated in Fig.3 of this paper and also in Hartley et al (2017), changes in the PFT fractional distribution exert significant impacts on the simulated water,

energy and carbon fluxes. I am not convinced that it's not important to differentiate bare soil from grassland in the PFTs. If the PFT product is intended for use only in model simulations with dynamic competition between PFTs, it needs to be stated explicitly in the paper.

Minor comments

Abstract, L16, 2D is not defined previously.

L95-96, I'd suggest to modify "this work aims to reduce the cross-walking component of uncertainty" to "this work aims to reduce the uncertainty in the cross-walking component"

L100, "...with existing high-resolution auxiliary data products that individually characterize one surface type with high accuracy." The authors need to provide the accuracy information of the auxiliary data products in Section 2.1 to support this argument, which are important for users to understand the uncertainties in the PFT product.

L152, "This CCI PFT product is based on v2.0.8 of the CCI MRLC time series", I can only find v2.0.7 data at <https://maps.elie.ucl.ac.be/CCI/viewer/>. Are v2.0.8 data available to users?

L230-235, Table A2 shows that there are small fractions for the shrub PFTs for classes 30-110, which seem to be in contradiction with the description here, i.e. "Pixels belonging to the shrubland classes (codes 120-122 and 180) can have a mixture of trees, shrubs, and herbaceous cover. For pixels of non-shrubland vegetation containing classes, the vegetated portion of the pixel is composed of trees and herbaceous cover". Can you explain?

L258-270, can you add the upper and/or lower limits in the text? They are not always included in the legend in Table 1.

L298-300, as I understand it, the sparse vegetation classes (150-153) may have some small trees but perhaps more likely to have shrubs than trees, especially if they are located above the tree line, please take a look at the Circumpolar Arctic Vegetation Map <https://www.caff.is/flora-cfg/circumpolar-arctic-vegetation-map>.

L357, "The bare area classes (codes 200, 201, and 202) can have up to 3 % vegetation cover, by definition", this vegetation cover information is not shown in Table 1. Can you

add such cover information (e.g. 3% etc.) mentioned throughout the paper in Table 1? So that it'd be easier for readers to understand the class codes and the definitions. In addition, I'd suggest to provide a reference.

L360-361, "the latter of which is estimated as 100 % minus the inland water percentage", this seems to be too high since the productivity of mosses and lichens is in general much lower than grasses. I'd suggest the authors to consult a LSM expert on this.

L379, $2^\circ \times 2^\circ$ is rather large, I wonder how many pixels are determined this way? I'd suggest to provide a percentage.

L410-411, "5) 96 % bare soil PFT and 4 % natural grass PFT (to meet the legend minimum of vegetation cover) are assigned to pixels of the sparse vegetation classes", should this be bare classes? Though previously described as "can have up to 3 % vegetation cover" instead of 4%.

Fig.1 (c), some needleleaved evergreen trees are distributed above the treeline, is this realistic? Are there field data or references to support this?

Fig.1(d) seems to show more coverage for Needleleaved deciduous trees than in the CCI Viewer and the tree cover map in Hansen et al (2013), can you explain why?

Fig. 1(g) and (h), there are large extent of needleleaved evergreen/deciduous shrubs, are there field data or literature to support this? I am not aware of the use of these PFTs in any models.

I'd suggest to use a scale bar with more levels, and perhaps the same scale bar can be used for the different PFTs maps in Fig.1.

L538, "grass vegetation may be assigned in some cases that might otherwise be a temporary bare area", can you elaborate a bit on this? How do you know that it might be "temporary bare area" vs. permanent bare area?

Section 4.2, it seems to me that there is not enough evidence to show that the new PFTs are more realistic than the previous ones. Thus it is hard to interpret results shown in Fig.5.

Table A2, note sum of fractions are either greater than 100% or <100% for some classes (e.g.10-40).