

Comment on **essd-2021-251**

Arthur Endsley (Referee)

Referee comment on "High-resolution land use and land cover dataset for regional climate modelling: a plant functional type map for Europe 2015" by Vanessa Reinhart et al., Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2021-251-RC2>, 2021

Overall, the paper seems to represent an advance in land-cover/ plant functional type (PFT) modeling that informs regional climate models (RCM). I think the paper could be improved by adding more detail about the cross-walk procedure and, in particular, providing a rich example of how the linkage to the Holdridge Life Zones (HLZ) allows for future updates to the land-cover classification in response to environmental change. I gather that, in an RCM, climate changes that push a location out of one HLZ and into another would prompt an update to the corresponding PFT at that location; i.e., the vegetation canopy would be allowed to change and therefore the RCM model state would change in advance of the next time step. This is very interesting and potentially powerful, given that it allows some of the complexity of a dynamic vegetation model to be represented more simply, but it should be made more explicit.

There are also a couple of key technical issues with the paper. The crosswalk procedure should be explained in much more detail; the tables in Appendix A are deceptively rich in detail but they really don't make sense without more explication in Section 3. I also think Figure 2 may have the colors for each HLZ mislabeled, as some of the map regions' displayed HLZ classes do not make sense (e.g., desert in Scotland). There are also additional analyses referred to by the authors (ca. Line 487, "The additional comparison with a high resolution dataset (WSF2015) showed that not only large but also small agglomerations of urban areas are represented well in LANDMATE PFT.") that are completely undescribed in the paper.

Specific comments:

Lines 12-13: The authors write: "A suitable evaluation method has been developed and applied to assess the quality of the new PFT dataset." This is a little vague. Perhaps the authors could be more specific about the type of evaluation method, or remove this sentence from the Abstract?

Lines 88-95 and Appendix A: I find the crosswalk tables in Appendix A, and the crosswalk procedure as described, to be confusing. The row sums for every crosswalk table in Appendix A are 100, which seems to imply they are percentages. However, using Table A4 as an example, it's not clear to me how the ESA-CCI LC class (number 30 in this example) is converted to the LANDMATE PFT based on the HLZ. Typically, a cross-walk table would describe how two uniquely identified, intersecting spatial units are resolved into mutually conflicting classifications. But neither the HLZ nor the ESA-CCI LC class are uniquely identified (i.e., two widely separated locations might have the same HLZ). Is this a probabilistic approach? If the pixel with ESA-CCI LC class 30 is in Holdridge Life Zones 1-6, it has a 20% chance of being classified as Tundra, 20 percent change of being classified as Swamp, and 60 percent chance of being classified as Crops? The authors should be much more explicit about this somewhere in Section 3.

Lines 93-94: The authors write: "This revision of the CWTs is supported by reference data and visual satellite image interpretation." Is this described somewhere? Please add a reference to the appropriate section of the paper (Section 3?).

Figure 2: I think the colors may not be labeled correctly... It seems that much of Italy and Spain are "Rain Forest" and "Wet Forest" variously in the "Warm Temperate" or "Boreal" climate zones while much of Scotland and Ireland appear to be "Desert scrub" or "Desert" (Warm temperate climate zone).

Section 3: I think the authors should add a table of their final PFT classification. It is confusing to have Table 3, which is specific to the validation dataset, but not a table of the PFT classification for this new data product. For instance, on Line 204, a reference is made to Table 3, which is the validation dataset's PFT table, during a discussion of the new product's PFT classification.

Line 174: Related to the previous comment... The authors write that the target set of PFT classes for trees and shrubs "was done at the expense of two shrub-PFTs." I interpret "at the expense of" as meaning that something was sacrificed, i.e., removed. I look at Figure 3 and see that there are, in fact, two shrub PFTs. Perhaps the authors could revise this sentence?

Line 315: I think the inline equation is missing an equal sign, i.e., the symbol " n_{i+} " should be followed by an equal sign.

Lines 487-488: The authors write: "The additional comparison with a high resolution dataset (WSF2015) showed that not only large but also small agglomerations of urban areas are represented well in LANDMATE PFT." This is the first time this analysis and this dataset (WSF2015) are mentioned. The authors should describe this analysis as part of the paper

(Section 4).

Technical corrections:

- Line 20: "was declared an" should be "were declared"
- Line 30: "as realistic" should be "as realistically"
- Line 44: "origin" should probably be "originate"
- Line 127: "are meraged" should be "are merged"
- Line 185: "lass 61" should probably be "Class 61"
- Line 216: A space is missing between the end of one sentence ("2005.") and the start of the next: "Although the ESA-CCI LC..."
- Line 428: I think Figure 7g needs to be reference along with Figure 7i.
- Line 489: "and despite of the" should probably be "and despite"