Review of EGUsphere-2026-3696

This manuscript presents an interesting approach to mapping trees on agricultural land by combining a sub-meter tree canopy map with different agricultural land-cover products. I consider the comparison of agricultural masks a valuable contribution, particularly because their effect on agroforestry estimates is not always evaluated. The application to Europe and the Sahel is also relevant, as these regions have very different agricultural systems. The use of high-resolution tree data is a strength of the study, especially for detecting scattered trees in dryland landscapes. Most of my comments relate to the definition of agroforestry, the validation data, the interpretation of accuracy metrics, and the uncertainty of the reported area estimates.

1. The abstract could better distinguish between the different accuracy metrics. The 77.2% reported for Esri (L21) is producer's accuracy, not overall accuracy. Also, the 45-percentage-point difference attributed to the agricultural mask compares products that differ in both cropland classification and pasture definition. It would be useful to clarify this. The statement that Esri performs best across all biomes is not fully supported by Fig. 4, where other products perform better in some cases.
2. Since the definition of agroforestry is central to this study, I think the manuscript could be clearer about what is actually being mapped. The method identifies trees or shrubs taller than 1 m on land classified as cropland or managed pasture. However, this does not necessarily mean that these trees are intentionally integrated into agricultural practices. Would "trees on agricultural land" be a more accurate description of the mapped product? This distinction is important when comparing the results with existing agroforestry estimates.
3. I think the choice of the 1 m tree-height threshold and 1% canopy-cover threshold needs more justification (L119–121, L162–163). For a 30 m pixel, 1% canopy cover corresponds to only about 9 m². The validation also considers a location positive when at least one tree pixel is detected. Could the authors show how the estimated areas change with different canopy thresholds, such as 1%, 5%, and 10%, and tree-height thresholds of 1 m and 2–3 m? This seems particularly important for the Sahel, where shrubs and small trees are common.
4. The use of the Human Footprint Index (HFI) to identify managed pastures needs further explanation (L140–146). HFI measures human pressure but does not directly indicate livestock grazing. In the Sahel, grazing is often mobile and may occur in areas with limited infrastructure or low population density. How well does the HFI > 4 threshold represent these systems? A sensitivity analysis using different thresholds would help assess its effect on the results. Please also clarify which Global Pasture Watch classes were included in the alternative masks.
5. The validation dataset is one of the strengths of the study, but I have some concerns about its composition. Approximately 7,000 of the 7,472 negative observations are forest points from Lesiv et al. (2022). These locations are likely to be excluded by the agricultural masks, which could partly explain the high specificity of 94–99%. In comparison, Esri correctly excludes only about 32% of woody monocrops. It would be useful to report specificity separately for each negative class and discuss how the composition of the reference dataset affects the overall accuracy results.

6. Related to the previous point, the validation would benefit from including more negative observations that are difficult to distinguish from agroforestry, such as trees on unmanaged grasslands, shrublands, orchards, vineyards, and olive groves. These seem particularly relevant for evaluating whether the maps can distinguish agroforestry from other tree-covered agricultural and semi-natural areas.
7. Please provide more information about the photo-interpreted reference dataset from Tahiri et al. (2026). Which imagery and acquisition dates were used? How many interpreters were involved, and was there any assessment of agreement between interpreters? These details would help readers understand the reliability of the reference data, particularly for classes such as scattered trees, alley cropping, and woody monocrops.
8. I think the manuscript could better separate errors associated with the tree canopy map from errors associated with the agricultural masks. In the current validation, a reference unit is considered mapped agroforestry if at least one tree pixel falls within the agricultural mask (L210–213). Could the authors report whether trees are detected independently of the agricultural classification? This would help determine whether omission errors result from missing trees or from agricultural land being classified as another land-cover type.
9. The accuracy results would benefit from confidence intervals and sample sizes, particularly for the biome-level comparisons in Fig. 4. Some differences in F1-score between products are relatively small. Since the same reference observations are used to evaluate all maps, a paired comparison could help determine whether these differences are meaningful. It would also be useful to clarify whether spatial clustering of reference observations was considered.
10. Since the area estimates focus on Europe and the Sahel, I think validation results should also be reported separately for these regions. How many reference observations are located in each region, and what are the corresponding accuracy metrics? Global validation results may not fully represent performance in these two agricultural environments.
11. One important methodological concern is the relationship between classification accuracy and area estimation. The Esri map is selected because it has relatively high producer's accuracy and F1-score, but these metrics do not necessarily indicate that it provides the most accurate area estimate. The reported 86 Mha for Europe and 36 Mha for the Sahel are obtained by counting mapped pixels, and their bias is not evaluated. I think this distinction needs to be made clearer in the methods and discussion. Skakun (2025) provides a useful discussion of this issue.
12. Related to the previous comment, could the authors consider a probability-based approach to estimate agroforestry area and its uncertainty, following Olofsson et al. (2014)? The existing map could be used to stratify a reference sample within each region. If this is beyond the scope of the current study, it is important to acknowledge that the reported areas are map-derived estimates and that their bias and uncertainty remain unknown.
13. I find the cropland and pasture results for the Sahel particularly interesting, but they also raise questions. The manuscript reports approximately 35 Mha of pasture with trees compared with only 1.17 Mha of cropland with trees (L336–338). Given the importance of rainfed agriculture and traditional agroforestry parklands in this region, could part of this difference result from cropland being classified as grassland or rangeland? It would be useful to compare the agricultural masks with an independent regional cropland dataset, or at least discuss this possibility and its implications for the cropland–pasture estimates.

14. The European results also deserve some additional attention. Figure 3 suggests that about 68% of woody monocrops are classified as agroforestry by the Esri-based map. Could this contribute to the large estimated extent in Mediterranean landscapes, where olive groves, orchards, vineyards, and other permanent crops are widespread? Reporting the overlap with available permanent-crop datasets would help clarify how these agricultural systems influence the European estimate.
15. I think the comparison with Liu et al. (2023) needs to distinguish between agricultural land containing trees and the area actually covered by tree canopy. The 86 Mha reported here represents agricultural land with at least some tree cover, while the 2.67 Mha from Liu et al. refers to a different quantity. Using the canopy-cover classes in Table 2 gives an approximate canopy area of 23 Mha in Europe, which considerably reduces the difference between the studies. The comparison would be more meaningful if the same quantities were used.
16. A similar issue applies to the comparison with Zomer et al. and Chapman et al. in Fig. 7. These products use different definitions, canopy thresholds, and spatial resolutions. Since the map developed in this study has not been independently established as a reference, describing other maps as overestimating or underestimating agroforestry may be too strong. I think these results should be presented as differences between products unless there is evidence that one estimate is more accurate.
17. The timing of the input datasets could be described more carefully. The agricultural masks cover different periods, the tree canopy map is based on imagery from 2017–2020, and some reference observations are from 2015. Agricultural land use and tree cover may have changed between these dates. Could the authors assess how temporal differences affect the validation, or explain their possible influence on the results?
18. L182: Please clarify which resampling method was used to bring the 10–30 m agricultural masks to 1 m. Resampling does not add spatial information to the original land-cover classifications. This is important because the resulting agroforestry maps still depend on the effective resolution of the agricultural masks, despite the finer resolution of the tree canopy data.
19. Please clarify the geographic boundaries used for Europe and the Sahel and provide the source of these boundaries. The European area reported in the manuscript suggests that European Russia may be included. This matters when comparing the results with studies focused only on the European Union.
20. L122–124: Please explain why protected-area categories Ia, Ib, and III were excluded, but category II was retained.
21. L133–137: The overall accuracies reported for Esri, ESA WorldCover, and Dynamic World come from different validation studies and reference datasets. These numbers should not be interpreted as a direct comparison of product performance without acknowledging these differences.
22. L169–171: Please clarify how the temporal mode was calculated for Dynamic World. Since Dynamic World provides frequent classifications while Esri provides annual maps, the two temporal aggregation procedures may not be equivalent. Also, Table 1 reports Esri coverage through 2023, while the methods mention 2024.
23. L239 and L350–351: Please explain how pixel area was handled during aggregation in Google Earth Engine. There are also small differences between the Sahelian canopy-cover areas reported in the text and Table 2, which should be corrected.
24. Figures 6 and 7: Please explain how the example locations in Fig. 6 were selected. Including randomly selected examples, or explaining the selection criteria, would help readers judge how

representative they are. In Fig. 7, please clarify whether the comparison uses Zomer et al. (2009) or Zomer et al. (2014), as the references appear inconsistent.

Overall, I think this study provides useful evidence that agricultural land-cover masks strongly affect the mapped distribution of trees on agricultural land. The high-resolution canopy data are an important advantage, particularly for dryland landscapes. However, the distinction between mapped area and independently validated area estimates needs more attention. Clarifying the definition of agroforestry, the limitations of the reference dataset, and the uncertainty of the regional estimates would make the results and conclusions more convincing.