Dear Reviewer,

Thank you for your constructive feedback. We provide our response below each comment in green font.

This paper describes the development of Copernicus Land Monitoring Service (CLMS) ETa data products and validation of the prototype product. I appreciate the authors' transparency in documenting these choices and opening the discussion on EGUsphere. However, the paper reads more like a combination of Algorithm Theoretical Basis Document and Product Quality Assessment Document, which should be published with CLMS data products anyway. Therefore, I'm questioning the scientific significance of this paper.
Although there is necessarily an overlap between some of the information content of this paper and ATBD and PQAD we do not think this detracts from the scientific contribution of this manuscript. The different documents have a different focus and while ATBD and PQAD are about informing the user of the product, this manuscript aims to focus on the scientific context of the product, e.g. through the content presented in Introduction and Discussion sections. For example, we believe that there is a clear scientific interest in exploring the selection and pre-processing of the various input data as the same inputs can be used in the other models. This is especially relevant since the data is free and open (coming mainly from Copernicus) and pre-processing methods are publicly available as open-source Python packages (see Code Availability section). Similarly, looking at the individual and ensemble outputs of just TSEB and ETLook models can lead to improved scientific understanding of the models themselves, but also to potential improvements in overall modelling of land surface water and energy fluxes.

General comments

1. I suggest the authors focus on specific research questions, rather than merely presenting the paper as "describing the design choices". Some lengthy details could be included in Annex to keep the focus on such research questions. For example, the paper could focus on investigating
   Perhaps a better phrase here would be "describing and justifying the design choices" since we not only described the methods used to prepare the various input data but also evaluate their performance and compare with alternatives (Section 4.3)

- Why the average of two models perform better than individual model?
- Where and when the ensemble average neutralizes biases of two models?
- What are the uncertainties and gaps in input data and processing that could influence model performance?

Those are all very valid suggestions and we will add a new section in the discussion where we analyze the performance of the two individual models and the ensemble in different plant functional types and climates and link that to the different model assumptions and limitations and uncertainties in input data.

2. The paper lacks a clear definition of ETa. You referred to Bojinski et al. (2014) for the definition of ECVs. However, in that article, "water use" was mentioned instead of "actual evapotranspiration". Table 1 shows that CLMS products do not include evaporation from canopy interception. Since interception contribution to total evaporation is substantial (even more than soil evaporation in many cases) (Savenije, 2004), and very an important process to hydrology (Dingman, 2015), I think it should not be neglected by CLMS.
The reviewer is right that water stored in the canopy by interception can sometimes be significant. However, some challenges are still open when estimating the evaporation in the canopy, such as i) consideration of not only rainfall but also dew, fog and overhead irrigation, which the latter ones are very difficult to use as inputs operationally, ii) there is no general model/framework for accounting for the amount of interception, LAI usually plays a role in such semi-empirical approaches, but also leaf size/shape as well as wind speed and rainfall intensity should be considered, furthermore iii) there is a lack of interception data that could be used to validate these models covering a wide range of canopies and climates, iv) the separation between soil evaporation, canopy transpiration and evaporation of stored water in the canopy is challenging (and in most cases missing) for in-situ reference EC datasets, and v) efforts to incorporate evaporation from intercepted water in the canopy, such as those implemented in GLEAM, ETMonitor or WaPOR assume that all available energy at the canopy is used to evaporate all intercepted water, without considering that in order for evaporation to happen the air must be dry enough (i.e. VPD also plays a role according to the Penman equation: $E = \frac{\Delta R_n + \rho_a c_p VPD}{\lambda(\Delta+\gamma) r_a}$). The latter issue implies that under a moist atmosphere, sensible heat flux can even be more significant than latent heat and thus temperature may increase in a canopy with intercepted water.

Furthermore, evaporation from canopy interception is implicitly integrated in the evaporation component on the energy balance, when using thermal infrared data as input, as this excess evaporation reduces the surface/canopy temperature. Should this process be considered robustly in energy balance models, a novel multi-source energy balance should be developed, which is not part of the scope of this manuscript.

3. Since canopy transpiration [mm/day] and soil evaporation [mm/day] are also CLMS products, effort should be made to evaluate their quality. You mentioned in L407 that these could not be contrasted against *in situ* However, there are references that could be used as proxy, e.g., partition of Eddy Covariance measurements (Nelson et al., 2020), and recently SAPFLUXNET (Poyatos et al., 2016). The ETLook model overestimates transpiration (compared to TSEB-PT) created large bias in ETa for specific cases, as you showed. The reason could also be that ETa from EC tower is underestimated due to energy balance closure (which was not discussed). In any case, investigating the accuracy of both models in transpiration estimation could add significant values in the discussion.

   Regarding the partition of evapotranspiration into its components, we agree that it would be an interesting exercise, however we do not think that it can be performed at a globally representative scale including the different climatic zones and plant functional types. The EC flux partitioning methods have rather high uncertainty, often require raw high frequency data (which is not usually available at higher level products of ICOS, Ameriflux and OxFlux sites) or site-specific data preprocessing and selection. Regarding the sap-flow measurements, there is the issue of spatial-scale mismatch when comparing against 300 m transpiration. For that reason, in the Product User Manual we specify that E and T partition is unvalidated and therefore has a larger uncertainty and we will add similar warning to the manuscript in the Discussion / Conclusion. We do however hope that once the dataset is public it will be used in various studies around the globe and the characterization of the partition will become more certain.

   Regarding energy balance closure we briefly discuss it on lines 411-420. However, in the revised manuscript we will make it more explicit. The validation results could potentially change depending on the energy balance closure (EBC) method and variants. In order to ensure transparency and minimize any scientific bias by us, we prioritize the use of the energy balance correction provided by the dataset, which usually is the Bowen Ratio method adopted by the FLUXNET community (Pastorello, 2020) and implemented in different datasets: ICOS, Ameriflux, OzFlux. In other stations we leave the flux uncorrected. How to deal with the EBC when validating ET models is still an open question for scientific debate.

   References:

   1. Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., et al. (2020). The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Sci Data 7, 225. https://doi.org/10.1038/s41597-020-0534-3.

4. The validation of the prototype is most limited in the tropical climate (Figure 5). The validation results also show lower performance than other regions. Meanwhile, the number of cloud-free images are lowest in this climate (Figure 2), even limiting eastern swath does not change much this fact. This points to a serious gap in a global data product like CLMS, which should be highlighted.
   Thank you for pointing this out. For the revised manuscript we are performing additional validation in South America, Asia and Africa and we will also highlight this limitation in the Discussion/Conclusion.

Specific comments

1. L14 “overall best” => “average” (of all site-dekads)
   What we meant by “overall best” was best out of TSEB-PT, ETLook and Ensemble. We will reword to make this clearer.
2. L16 which are the similar global ETa dataset mentioned?
   It is the WaPOR dataset. We will also make this clearer.
3. L21 What is the definition of “operational” in this context? Does it mean “near-real time monitoring” or something else?
   Copernicus definition of operational is (paraphrased) produced at regular intervals in near-real-time using well established and validated methods and input data sources and with guaranteed long-term continuity. It also means that the products largely satisfy the applicable requirements, that their limitations are non-critical and documented, and that regular quality monitoring and assessments are performed to check and ensure the stability of product quality. We will define it properly in the revised manuscript.
4. L45-47 Could you provide some references?
   This is based on official and unofficial discussions with FAO and therefore does not have references. FAO is aware of this manuscript.
5. L48 What are the other global CLMS products? What are their spatial and temporal resolution?
   Some example of those products are top-of-canopy reflectance (300 m, daily), vegetation products (Leaf Area Index, Fraction of Absorbed Photosynthetically Active Radiation, NDVI - 300 m, dekadal) and primary production products (GPP, NPP, Gross Dry Matter Productivity - 300 m, dekadal) - https://land.copernicus.eu/en/products/vegetation. We will mention this in the revised manuscript.

6. L48-L59 This paragraph seems to justify the spatial and temporal resolution of CLMS (300m, 10-d). If continuing WaPORv3L1 is the main factor, it should be expressed more explicitly. Other products can be presented as reference for state-of-the-art.
   In this paragraph we present examples of ET datasets (known to us) which are produced operationally and in near-real-time with a global extent or satisfy at least some of those conditions. This is done to place CLMS ET product in the context of other, similar products and not to justify the spatial and temporal resolutions. Those resolutions were determined (requested) by European Commission and were the constraints within which we conducted the study and were based on a number of factors:
   - Other CLMS products have the same spatial and/or temporal resolutions.
   - 300 m is the spatial resolution of the Sentinel-3 OLCI sensor
   - The manuscript describes design of a global product which aims to have long-term continuity. Increasing the spatial or temporal resolutions would have an impact on the data storage and distribution costs.

   The fact that WaPORv3L1 and CLMS spatio-temporal resolutions align is a coincidence or could point to WaPOR being influenced by existing CLMS products at the time its L1 product was being defined.
   We will improve the justification for the spatial and temporal resolutions in the revised manuscript.

7. L60 What are these preparatory activities? Who recommended two ET modelling frameworks? Why were these two selected?
   We will modify the paragraph on lines 60-65 to the following:
   Preparatory activities conducted by European Commission to develop an operational CLMS ETa product recommended that two ET modelling frameworks should be further investigated. The first one is the Sen-ET framework (Guzinski et al., 2020, 2021) developed to model ETa with Copernicus data at various spatial scales and using the Two-Source Energy Balance Priestley-Taylor (TSEB-PT) ET model (Norman et al., 1995; Kustas and Norman, 1999; Anderson et al., 2024). The second is the WaPOR framework (FRAME Consortium & FAO, 2024) developed by FAO through the WaPOR project and using the ETLook ETa model (Bastiaanssen et al., 2012). Both models, although conceptually different, estimate evaporation and transpiration and use LST as one of core input forcings. This recommendation was mainly based on the availability of mature open-source implementations of the two ET models, on previous studies demonstrating the applicability of both models with Copernicus data sources (i.e. Sentinel-3 imagery and meteorological data from European Center for Medium Range Weather Forecasts) (Guzinski et al., 2025) and on FAO's familiarity with both approaches. This does not imply

that those two modelling frameworks clearly outperform all other approaches and indeed it has been shown that the performance of an individual model depends on the landcover and climatic conditions (Reitz et al, 2025). However, due to constraints on developing a publicly funded global and operational dataset (outlined previously) there was a need to limit the design of the first version of the CLMS ETa product to those two frameworks.

References:

1. Guzinski, R., Nieto, H., Sandholt, I., and Karamitilios, G. (2020). Modelling High-Resolution Actual Evapotranspiration through Sentinel-2 and Sentinel-3 Data Fusion. Remote Sensing 12, 1433. https://doi.org/10.3390/rs12091433.

2. Guzinski, R., Nieto, H., Sánchez, J.M., López-Urrea, R., Boujnah, D.M., and Boulet, G. (2021). Utility of Copernicus-Based Inputs for Actual Evapotranspiration Modeling in Support of Sustainable Water Use in Agriculture. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14, 11466–11484. https://doi.org/10.1109/JSTARS.2021.3122573.

3. Norman, J.M., Kustas, W.P., and Humes, K.S. (1995). Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature. Agricultural and Forest Meteorology 77, 263–293. https://doi.org/10.1016/0168-1923(95)02265-Y

4. Kustas, W.P., and Norman, J.M. (1999). Evaluation of soil and vegetation heat flux predictions using a simple two-source model with radiometric temperatures for partial canopy cover. Agricultural and Forest Meteorology 94, 13–29. https://doi.org/10.1016/S0168-1923(99)00005-2.

5. FRAME Consortium & FAO: WaPOR methodology, https://github.com/un-fao/wapor-et-look/wiki, 2024.

6. Bastiaanssen, W.G.M., Cheema, M.J.M., Immerzeel, W.W., Miltenburg, I.J., and Pelgrum, H. (2012). Surface energy balance and actual evapotranspiration of the transboundary Indus Basin estimated from satellite measurements and the ETLook model. Water Resources Research 48. https://doi.org/10.1029/2011WR010482.

7. Reitz, M., Volk, J.M., Ott, T., Anderson, M., Senay, G.B., Melton, F., Kilic, A., Allen, R., Fisher, J.B., Ruhoff, A., et al. (2025). Performance Mapping and Weighting for the Evapotranspiration Models of the OpenET Ensemble. Water Resources Research 61, e2024WR038899. https://doi.org/10.1029/2024WR038899.

8. L66 Related to my first general comment, if the paper only “aims to give an overview of the design choices”, it could be just a technical documentation. To make a clearer scientific significance, this paper should focus on investigating further the two model frameworks (L61).

We agree and will therefore add a section in the discussion focusing on the two modelling frameworks (as mentioned previously). However, we would again like to highlight that we believe that describing, and justifying, the forcing pre-processing step of a global ET product has scientific value in itself.

9. Section 2. I would start with 2.8 first to provide an introduction about the two modelling frameworks, clearly indicate which input variables are required for both/each framework. Then continue with 2.1 and link required variables with input data sources in Table 2, indicating which one is used in which model.
   Thank you for this suggestion, we will reformat this section as suggested.

10. Table 2. How were 100m wind speed, air temperature, and water vapour pressure converted to near surface level (e.g., 2m is commonly used).
    In our model setup we use the wind speed, air temperature and water vapour at 100 m above the surface. Due to the low spatial resolution of those fields we assumed that they represent conditions at atmospheric blending height (set to be 100 m above ground) where impact of local surface conditions on those parameters is not so direct (Guzinski et al., 2021).
    References:
    1. Guzinski, R., Nieto, H., Sánchez, J.M., López-Urrea, R., Boujnah, D.M., and Boulet, G. (2021). Utility of Copernicus-Based Inputs for Actual Evapotranspiration Modeling in Support of Sustainable Water Use in Agriculture. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14, 11466–11484. https://doi.org/10.1109/JSTARS.2021.3122573.

11. L81 Is this in the mandate of CLMS?
    The mandate of CLMS and Copernicus is (from lines 81-82) “to ensure the free and open license conditions, long-term future continuity and consistency across the CLMS portfolio”. One way of achieving this is to base input data predominantly on Copernicus data sources.

12. If Copernicus data product should be based on Copernicus data, why the canopy height map was derived from a static map by ETH Zurich and not derived from other Copernicus data dynamically (e.g., Sentinel-2)?
    The ETH Zurich product is based on GEDI LiDAR measurements combined with Sentinel-2 imagery. There is no Copernicus equivalent of GEDI and, therefore, in this case there was no alternative. However, the tree canopy height is an ancillary product that remains largely static (in a global perspective) and we further filter it through the CLMS landcover product

(lines 313-317). We should clarify thus that canopy for herbaceous and annual canopies are dynamic and estimated based on LAI.

13. Table 3. The native resolution of these input data sources should be included.
    Thank you for the suggestion. We will add this information to the table.
14. L104 missing reference for semi-Bayesian approach
    The reference will be added:
    Bulgin, C.E., Sembhi, H., Ghent, D., Remedios, J., & Merchant, C.J. (2014). Cloud Clearing Techniques over Land for Land Surface Temperature Retrieval from the Advanced Along Track Scanning Radiometer. International Journal of Remote Sensing, 35, 3594-3615
15. L108 What if LAI changes in case of crop harvesting, forest fire?
    All gap-filling approaches have their limitations and this is certainly one of them. But in large majority of cases vegetation follows the phenological cycle and this is frequently used during gap-filling (e.g. the popular TIMESAT method or Whittaker smoother used in WaPOR).
16. L109 Can the gaps in LAI be suitable for filling in areas where it is cloudy most of the year?
    Similarly to previous answer, gap-filling over long gaps or with few data points is for sure challenging and in general the data quality decreases with longer gap periods.
17. L112 needs reference
    We will add following reference:
    Abbasi, B., Qin, Z., Du, W., Li, S., Fan, J., and Zhao, S. (2020). Effects of Cloud on Land Surface Temperature (LST) Change in Thermal Infrared Remote Sensing Images: a Case Study of Landsat 8 Data. In IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium (IEEE), pp. 5430–5433. https://doi.org/10.1109/IGARSS39084.2020.9324415.
18. L185 needs references. Not clear to what degree of accuracy, and from which EO data.
    We will rephrase this sentence. We just wanted to stress that both albedo emissivity and LST are routinely estimated with EO data, but we should never forget that even these rather mature products are not free for uncertainties as any other remote sensing variable
19. L187 How do these models differ in Ln calculation?
    In the TSEB-PT model, Ln is computed according to equation 2a from Kustas & Norman (1999) using CAMS surface_thermal_radiation_downwards field together with internally computed leaf and soil temperatures, constant emissivity values for the vegetation and soil (0.98 and 0.95 respectively) and effective leaf area index. WaPOR-ETLook estimates it

following the FAO-56 approach and daily values of air temperature, vapour pressure and transmissivity (FRAME Consortium & FAO, 2024).

References:

1. Kustas, W.P., and Norman, J.M. (1999). Evaluation of soil and vegetation heat flux predictions using a simple two-source model with radiometric temperatures for partial canopy cover. Agricultural and Forest Meteorology 94, 13–29. https://doi.org/10.1016/S0168-1923(99)00005-2.

2. FRAME Consortium & FAO: WaPOR methodology, https://github.com/un-fao/wapor-et-look/wiki, 2024.

20. L2555 needs reference for “most approaches”

    Here we refer to DMS, TsHARP and other similar methods (see e.g. Gao et al., 2012; Agam et al., 2007; Sanchez et al, 2020). We will add those references to the revised manuscript.

    References:

1. Gao, F., Kustas, W.P., and Anderson, M.C. (2012). A Data Mining Approach for Sharpening Thermal Satellite Imagery over Land. Remote Sensing 4, 3287–3319. https://doi.org/10.3390/rs4113287.

2. Agam, N., Kustas, W.P., Anderson, M.C., Li, F., and Neale, C.M.U. (2007). A vegetation index based technique for spatial sharpening of thermal imagery. Remote Sensing of Environment 107, 545–558. https://doi.org/10.1016/j.rse.2006.10.006.

3. Sánchez, J.M., Galve, J.M., González-Piqueras, J., López-Urrea, R., Niclòs, R., and Calera, A. (2020). Monitoring 10-m LST from the Combination MODIS/Sentinel-2, Validation in a High Contrast Semi-Arid Agroecosystem. Remote Sensing 12, 1453. https://doi.org/10.3390/rs12091453.

21. L269 needs reference

    We will add the missing reference:

    FRAME Consortium & FAO: WaPOR methodology, https://github.com/un-fao/wapor-et-look/wiki, 2024.

22. L276 Is this statement based on sensitivity analysis? I would expect aspect and slope influence shaded relief (related to shadow effect, intensity of sunlight exposure), which influence LST. (e.g., Peng et al., 2020)

    This is true, but this is already account for in the “cosine solar inclination angle” parameter which is derived from aspect and slope. We will add this clarification in the revised manuscript. This is backed up by analysis shown in section 4.3.2.

23. L291 Only topographic correction?
    Yes, only topographic correction of the meteorological variables shown in Table 6 was performed. However, this results in meteorological forcings suitable for evapotranspiration modelling as shown in Figure 22.
24. L313 resampled by averaging?
    Yes by averaging, we will make this clear in the revised manuscript.
25. L398 How close is the closest non-gap-filled dates? Do you apply a threshold for when the date is too far?
    The closest gap-filled date varies by location and season and can range from 1-day to 60-days (line 394).
26. L416 Although it is mentioned here that you searched for more diversity in the reference datasets, only Eddy Covariance methods are considered. Meanwhile, many other references have been used to (qualitatively) validate remotely sensed ET, especially in regions lacking Eddy Covariance measurements (Tran et al., 2023)
    Please see reply to comment number 27.
27. For example, a large area in Figure 4 has no stations at all. These areas would require alternative effort and methodology to evaluate CLMS data products. For examples, see de Andrade et al. (2024) for South America, see Weerasinghe et al. (2020) and Blatchford et al., (2020) for Africa, see Athira et al. (2025) for India, see He et al. (2025) Southeast Asia, see Cogill et al. (2025) for South Africa. I think at least, the authors should refer to these studies for future validation of CLMS data products.
    Thank you for the reference. It is true that many areas do not have Eddy Covariance (EC) measurements but the focus of qualitative validation was specifically on this data as it is the most widely used for validation of satellite-based ET estimates. Also please note that in the revised manuscript, we will add extra EC sites in South America, Africa and Asia. In addition, we will mentioned the suggested references when discussing the limits of validation in the revised manuscript.
28. L440 Which quality flag was excluded? And what was considered realistic? Based on what?
    No quality flag was excluded. On the contrary, the quality information provided with each dataset was effectively used to ensure the selection of good quality data. The point on realistic range refers to cases in which missing data are filled with numerical values (*e.g. -999.9*) that are meant to indicate that the timeslot does not contain a valid value. The occurrence of such values goes normally together with a bad quality score in the quality

flag but using both approaches gave an additional warranty of good data selection. We will clarify this in the revised manuscript.

29. L446 What I missed here is whether ETa values were extracted at the 300-m pixel containing the station or flux footprint was considered. The mismatch in spatial support although not affect overall results but would affect some cases where the flux site is located in heterogenous area (e.g., IT-BCi).
    The ETa values were extracted from the 300m pixels containing the station. Evapotranspiration validation is performed during the day, therefore under unstable conditions. In these cases, the footprint is generally limited to a few hundred meters or less. We will add this information to the revised manuscript. However, spatial scale mismatch can indeed affect some sites and this will be mentioned in the revised manuscript.
30. Table 8. rBias should also be included for site level
    Thank you for this suggestion. Table 8 will be modified accordingly in the new version of the manuscript.
31. L462 The overall statistics in Appendix C would not change much, but in my experience, bias and rbias would differ remarkably for some sites that energy balance disclosure is high.
    Yes this is true, here we were focusing on the overall statistics and we will mention this in the revised manuscript.
32. L481 "better than" what?
    Better than individual models. This will be clarified in the revised manuscript.
33. L483 what do you mean by "the variability of the fluxes"? temporal or spatial variability?
    We mean the standard deviation shown in Taylor plots of Figure 9.
34. L484 reference to Figure 7?
    Again here we refer to the standard deviation shown in the SAV plot of Figure 9, where standard deviation of ETLook values is larger than in situ (variability is overestimated) while standard deviation of TSEB-PT is smaller than in situ (variability is underestimated). We will improve the clarity of this in the revised manuscript.
35. L486 Interesting. I think this should be investigated further in future development.
    We agree that this should be investigated further since forests are generally complex ecosystems for which it is most difficult to model ET. Please note that the data which will

be used in the updated manuscript will be derived with TSEB-PT and ETLook models in which parameterization was updated and this is expected to reduce the magnitude of this difference.

36. Figure 10. Could you also plot the EB-corrected data for tower? What does the error band around "Tower" line represent?
    The line with the label 'Tower' in that Figure is actually the EB-corrected data. That clarification will be added to the caption of the Figure in the revised version of the manuscript. The band around the line is the random uncertainty. The random uncertainty –when provided in the eddy covariance dataset- for a particular timeslot is derived from flux measurements with similar meteorological conditions within a time window of +/- 5 days around the date and within a window of +/- 1 hour around the time of the day, as described in Pastorello et al, 2020.
    Reference:
    1. Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., et al. (2020). The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Sci Data 7, 225. https://doi.org/10.1038/s41597-020-0534-3.

37. L491, ETLook is not visible in the Taylor plot (Fig9-EBF).
    The figure will be corrected in the revised manuscript.

38. L496 Here it says the ensemble formula is a suitable option in the case of DBF (Figure 11). But in the case of EBF (Figure 10- IT-Cp2), the ensemble is worse than TSEB-PT, so would you say the ensemble formula is not applicable everywhere?
    Yes we agree that ensemble does not perform better than the individual models in all sites. However, overall it still presents the most accurate results. It should be noted that in the CLMS ET product the user will have access to ET produced by ensemble and the individual models. We will add this information to the revised manuscript.

39. L525 The description of WaPOR global ETa dataset should be mentioned before the validation results (Section 3)
    We did not consider comparison with WaPOR to be validation which is why this is only described in the Discussion. However, we decided to include WaPOR ET in the plots and tables of Section 3.2 not to duplicate the figures and to make the comparison easier. We agree that this might be confusing and therefore we will make a reference and brief description of WaPOR ET before Section 3.2 but discuss the comparison still in Section 4 or we will remove WaPOR results from the Tables and Figures presented in section 3.

40. L530 a low bias presents high accuracy, so I would say “performance” instead of “accuracy” here.
We will correct this in the revised manuscript.

41. L566 The selection of tiles seems arbitrary to me. I would suggest more strategic selection, focusing on tiles without flux sites, tiles with high heterogeneity (e.g., indicated by coefficient of variation of vegetation index), and tiles represent climatic zones (especially Tropical and Arid, since these are lacking validation sites).
Thank you for the suggestion, we will consider it when updating the figure for the revised manuscript.

42. Figure 15. Although it might make the histogram look crowded, I’d suggest including the ensemble.
Again thank you for the suggestion, we will take it into consideration.

43. Figure 17. What does ETA stand for?
On the y-axis are LAI, gLAI and fg derived with the CLMS LAI/FAPAR (labeled as CLMS) dataset and on the X from the biophysical model produced as part of the CLMS ET processing (labeled as ETA). We agree that this might be confusing and it will be corrected in the revised manuscript.

44. L618-619 This explanation is not clear to me. The uncertainty in LAI at higher LAI should affect ETLook model more since LAI is used to partition available energy for transpiration and soil evaporation in ETLook. I think saying that it has “minimal effect on ETa modelling” requires some local sensitivity analysis or uncertainty analysis.
LAI also affects the partitioning between E and T in the TSEB-PT model, in particular LAI is also used in the Campbell radiative transfer model to partition the (net) radiation into canopy and soil (see Eqs. 3a and 3b). Indeed, at high LAI values most of the gap fraction becomes negligible according to the exponential law of Beer-Lambert, and thus most of the partitioning is already assigned to the canopy and most of the water flux to T. Therefore, a change of LAI between e.g. 4 and 5 does not significantly affect this partitioning.

45. L628-630 I’m connecting this larger bias in LAI in dense vegetation (e.g., broadleaf forest) with the lower performance of ETLook in EBF and DBF discussed in L486-489. Maybe a simple local sensitivity analysis could help address this.
This could be part of the reason but LAI is equally important parameter for TSEB-PT model (see the answer above). We have recently discovered some parameterization issues in our implementation of ETLook which we think could be major contributors to this lower

performance. In the revised manuscript we will use data produced with updated parameterization.

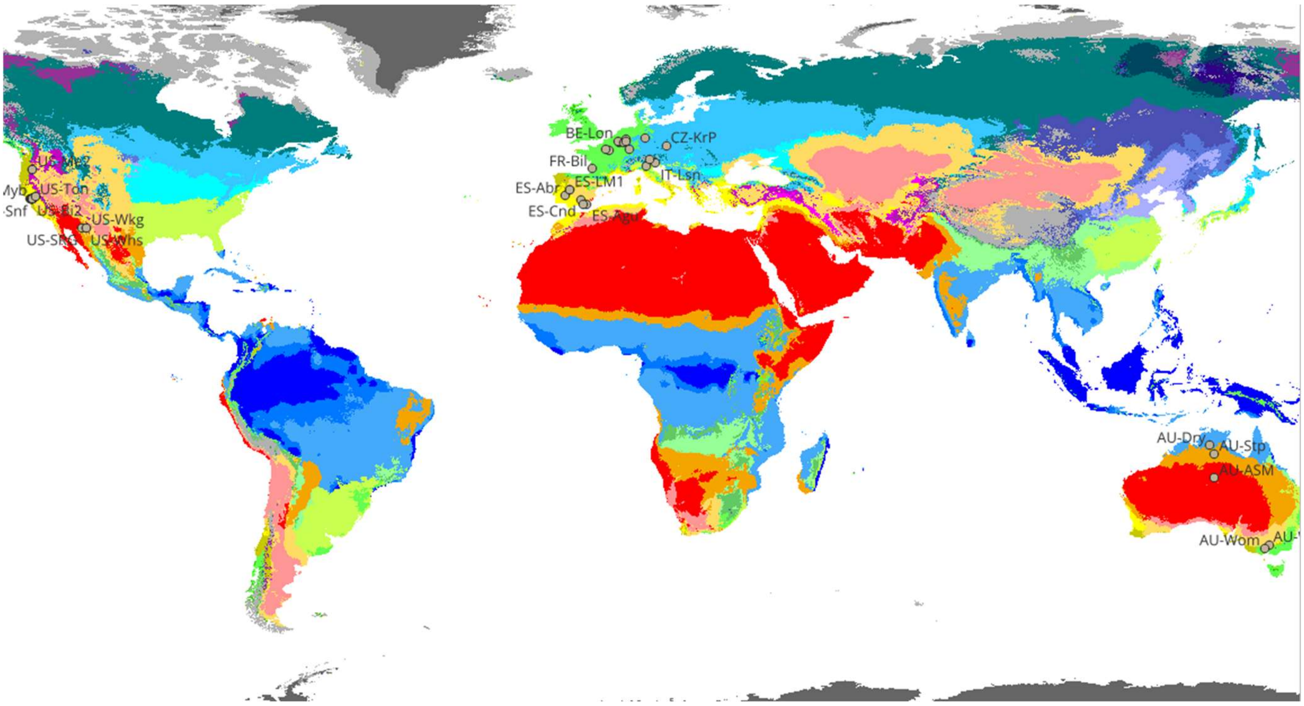
46. L633-656 Similarly, the inconsistency of CLMS LAI at SAV site could contribute to why the r2 of ETLook is lower than 0.5 for SAV (Figure 7)
    Please see the answer above.

47. L660 How were these AOIs selected? Based on which criteria? Why not selecting AOI based on level of heterogeneity/homogeneity?
    This was part of the selection process, as was the aim to cover different climatic zones and ecosystems distributed around the globe. We will add this information to the revised manuscript

48. L672-L680 The site-level results were discussed but not presented or referenced.
    The table with site-level results was not included in the manuscript for brevity but will now be added to supplementary material.

49. L690 Interesting. Could you show some examples? Where were these artifacts commonly observed? I think this would be important for users to know since the use of products like CLMS are often for spatial analysis.
    We observed them e.g. in tropical rainforest with uniform dense vegetation. In the revised manuscript we will show an example in the Supplementary Material.

50. L696-L700 Do these 45 sites mainly in Europe also represent Tropical and coastal regions?
    We will make a correction since there were 43 sites, but there are 3 sites in Europe closer than 25km to the Mediterranean (ES-Agu, ES-LJu, IT-Lsn), and 4 in US close to the Pacific (US-CGG, US-Myb, US-Sne and US-Snf). However, only one site is on tropical region AU-Dry (Aw Köppen climate)



51. Figure 22. Could you say something about the cluster of points with EC below 100 Wm-2 but spread from 100 to 350 Wm-3 for ETA?
    Those points most probably represent situations with persistent (on a daily time-scale) local cloud/fog which was not representative of the conditions seen in the CAMS pixels with its 0.4° spatial resolution. However, it could also be due to dust, spider webs or other obstructions on the upward pointing radiation sensors mounted on the EC tower.

52. Overall Section 4.3 presents a lot of methodological details and could be more focused on the impact on the prototype products. I would suggest methods or trials of methods to be described in Methodology section or Annex.
    In our opinion the trials of methods present some of the largest scientific input of this paper (apart from the ETa product itself), since they (and the lessons learned) can be applied to other global and regional products. This is especially true since the core of those methods are all open-source as listed in the "Code availability" at the end of the manuscript. However, we also agree that more emphasis should be placed on evaluating the strengths and weaknesses of the two models and the ensemble and therefore we will add a dedicated section on this topic to the discussion in the revised manuscript.

53. L733 ECOSTRESS could also be considered, then.
    ECOSTRESS is for sure a very useful sensor for prototyping of different methods. However, with its irregular acquisition times, frequency and coverage of the globe and with its

limited duration on board of the International Space Station it is not suitable for global operational applications focused on long-term continuity.

54. L747 Also canopy interception
    As mentioned in the response to comment number 1, interception is already implicitly captured in the evaporation component. Explicitly separating evaporation from canopy interception is scientifically challenging, however we will add it to potential improvements in the revised manuscript.

55. L761 Could you say that the pre-processing methods are "applicable globally" when the evaluation in Tropical climate (areas with least validation sites and most data gaps) was not included?
    The selected methods rely on globally available data sources and do not undergo any localized fine-tuning or specific parameterization. In that sense they are applicable globally. However, we agree that the quality of the outputs is not well characterized globally and the models might perform better in some areas than others. We will add such statement to the revised manuscript. We would also like to point out that some stations in the tropical climate were already used (northern Australia and Indonesia) and in the revised manuscript we will use additional stations from South America, Africa and Asia.

References

Athira, K.V., Rajasekaran, E., Boulet, G., Nigam, R. and Bhattacharya, B.K., 2025. Multiscale evaluation of three remote sensing-based evapotranspiration models across the humid to arid tropics: a study over India. *Hydrological Sciences Journal*, pp.1-22.

Blatchford, M.L., Mannaerts, C.M., Njuki, S.M., Nouri, H., Zeng, Y., Pelgrum, H., Wonink, S. and Karimi, P., 2020. Evaluation of WaPOR V2 evapotranspiration products across Africa. *Hydrological processes*, *34*(15), pp.3200-3221.

Cogill, L.S., Toucher, M., Wolski, P., Esler, K.J. and Rebelo, A.J., 2025. Evaluating the performance of satellite-derived evapotranspiration products across varying bioclimates in South Africa. *Remote Sensing Applications: Society and Environment*, p.101612.

de Andrade, Bruno Comini, Leonardo Laipelt, Ayan Fleischmann, Justin Huntington, Charles Morton, Forrest Melton, Tyler Erickson et al. "geeSEBAL-MODIS: Continental-scale evapotranspiration based on the surface energy balance for South America." *ISPRS Journal of Photogrammetry and Remote Sensing* 207 (2024): 141-163.

Dingman, S.L., 2015. *Physical hydrology*. Waveland press.

He, S., Yang, Q., Zhang, L., Shi, Z., Wang, X. and Lv, B., 2025. Consistency Assessment and Uncertainty Analysis of Spatial-temporal Characteristics of Evaporation Data in the Greater Mekong Subregion. *Journal of Hydrometeorology*.

Nelson, J.A., Pérez-Priego, O., Zhou, S., Poyatos, R., Zhang, Y., Blanken, P.D., Gimeno, T.E., Wohlfahrt, G., Desai, A.R., Gioli, B. and Limousin, J.M., 2020. Ecosystem transpiration and evaporation: Insights from three water flux partitioning methods across FLUXNET sites. *Global change biology*, *26*(12), pp.6916-6930.

Peng, X., Wu, W., Zheng, Y., Sun, J., Hu, T. and Wang, P., 2020. Correlation analysis of land surface temperature and topographic elements in Hangzhou, China. *Scientific Reports*, *10*(1), p.10451.

Poyatos, R., Granda, V., Molowny-Horas, R., Mencuccini, M., Steppe, K. and Martínez-Vilalta, J., 2016. SAPFLUXNET: towards a global database of sap flow measurements. *Tree physiology*, *36*(12), pp.1449-1455.

Savenije, H.H., 2004. The importance of interception and why we should delete the term evapotranspiration from our vocabulary.

Tran, B.N., Van Der Kwast, J., Seyoum, S., Uijlenhoet, R., Jewitt, G. and Mul, M., 2023. Uncertainty assessment of satellite remote-sensing-based evapotranspiration estimates: a systematic review of methods and gaps. *Hydrology and Earth System Sciences*, *27*(24), pp.4505-4528.

Weerasinghe, I., Bastiaanssen, W., Mul, M., Jia, L. and Van Griensven, A., 2020. Can we trust remote sensing evapotranspiration products over Africa?. *Hydrology and Earth System Sciences*, *24*(3), pp.1565-1586.