Dear Annemarie,

Thank you for your positive review and valuable suggestions to improve the manuscript. . We provide our response below each comment in green font.

GENERAL COMMENTS

Congrats on your work. I was involved in the development of the WaPOR v3 database and I am well aware of the challenges you faced. Impressive work. The paper is clearly based on years of experience in the topic, and presents an overall thorough research.

My recommendation is to consider the paper for publication although some revisions are required.

MAJOR COMMENTS

Gapfilling procedure: Section "2.9 Output gap-filling" describes the procedure used to fill gaps. It is chosen to fill gaps in the outputs and not in the inputs with the argumentation that the satellite observations (e.g. LST, LAI, albedo) are all acquired at the same moment, and will have the same gaps. The output gaps are filled using KcKs*RET. Ks can change daily or even within a day, especially when soil moisture is depleted, or rainfall/irrigation happens. You are likely to overestimate actual ET as Ks is reduced under cloudy conditions. Ks is likely to be higher during cloud-free periods when plant have higher water demands. The problem with this method is that KS is not a fixed crop property, but depends on soil moisture and atmospheric demand. It would be more logical to gapfill soil moisture, as this is more constant over time, and you preserve the physical relationship. I do not think the corrections for rainfall are sufficient to overcome this weakness. Adding to this, it is not clear how the KcKS method was used to create decadal data. Did you calculate an Eta value for each day, or did you create decadal KcKs values?
We agree that no gap-filling method is perfect and using $K_{c,s}$*RET also has its limitations. Nevertheless, it is a widely used method to gap-fill actual evapotranspiration estimates (including in the Sen-ET framework) and is of comparable skill to other widely used gap-filling methods (see e.g. Delogu et al., 2021). A large part of the changes in atmospheric demand are already captured in the RET calculation and we try to minimize the effect of soil moisture by using $K_{c,s}$ derived as closely as possible to the target date. We further try to account for changes in SM by e.g. taking rainfall into account. Gap-filling soil moisture might work for ETLook (but not for TSEB which does not use it as input) but itself contains limitations. For example, the Whittaker smoother used in the WaPOR framework assumes smooth variability in SM which might not be necessary true during rainfall, irrigation or drought events.
Regarding the creation of dekadal values, we will add following sentence to the last paragraph of

section 2.9 (line 398):
"The all-sky gap-filled daily estimates of evapotranspiration, evaporation and transpiration are then aggregated to dekadal timesteps by taking the mean of all valid values within the aggregation period."
References:
1. Delogu, E., Olioso, A., Alliès, A., Demarty, J., and Boulet, G. (2021). Evaluation of Multiple Methods for the Production of Continuous Evapotranspiration Estimates from TIR Remote Sensing. Remote Sensing 13, 1086. https://doi.org/10.3390/rs13061086.

Ensemble: The suggestion for an Ensemble model does miss a proper defense, where are the complementary strengths of the models? When one model is overestimating, and the other underestimating, the ensemble may appear closer to observation, but not necessarily because it captures the underlying processes better, it is simple averaging out the opposite biases.
Ensemble models (or ensembles of model outputs) often perform better than individual models due to cancellation of random errors, and sometimes bias, produced by the individual models (Volk et al. 2024). Indeed this technique is widely applied in weather forecasting, climate change predictions and hydrological modelling. This does not necessarily mean that they better capture the underlying processes. In the revised manuscript we will add a new section to the discussion in which we analyze the assumptions and limitations of the two models through their performance in different climates and plant functional types and which might help to address this issue. We will also add a paragraph in section 4.4 (Potential improvements) on a more robust method for producing ensemble which takes the model performance into account based on the recent experience of the openET consortium (Reitz et al, 2025).
References:
1. Volk, J.M., Huntington, J.L., Melton, F.S., Allen, R., Anderson, M., Fisher, J.B., Kilic, A., Ruhoff, A., Senay, G.B., Minor, B., et al. (2024). Assessing the accuracy of OpenET satellite-based evapotranspiration data to support water resource and land management applications. Nat Water 2, 193–205. https://doi.org/10.1038/s44221-023-00181-7.
2. 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.

Imbalance in explaining design choices: The paper does show an imbalance in methodological detail that affects understanding the key design choices. PROSPECT modelling is described exhaustively while other critical decisions receive less attention:

Gapfilling approach (see above)

Due to the already lengthy nature of this manuscript, we tried to focus on areas which were not previously described in detail in other papers. For topics which were described in previous publications, such as the two ET models, gap-filling or thermal sharpening, we provide a shorter description and references to those other studies.

“The CLMS ETa product specification states a spatial resolution of 300 m. However, the spatial resolution of the SLSTR LST product is 1 km.” I understand that CLMS has a strong preference for Copernicus datasets, but I do miss an explanation on why Sentintel-3 LST at 1km has been chosen instead of higher spatial resolution datasets such as VIIRS, and what is the impact of this decision, except in section 4.4.

The use of Sentinel-3 LST was one of the constraints imposed on the CLMS ET product by the European Commission. The major reasons for this are mentioned on lines 81-82 (“This is to ensure the free and open license conditions, long-term future continuity and consistency across the CLMS portfolio”). In particular, the long-term continuity of VIIRs is not guaranteed. Since this was a constraint within which the study was conducted, an analysis of potential impacts on this decision was not performed. However, we briefly mention a potential impact when discussing the differences between WaPOR and CLMS ETLook outputs in section 4.1. In the revised manuscript we will add a paragraph to the introduction explaining those constraints.

CAMS data processing – needs clarification, also on how historical data is derived.

The CAMS data is used for both NRT and historical production, in both cases using the shortest possible forecast duration of 12 hours. We will add this clarification to section 2.6. The reason for this is that the first version of CLMS ET product will be produced only in the near-real-time mode, e.g. there will be no reprocessing 2-3 months after the production date to perform improved gap-filling or use reanalysis meteorological data (like is done e.g. in WaPOR). Therefore, using data source other than CAMS for data produced before November 2025 (when operational production started) and CAMS for data produced after that date could lead to discontinuity of the timeseries. In Figure 22 we demonstrate that the CAMS data is highly suitable for use in this application. Nevertheless we also mention the production of reanalysis dataset, which could use ERA 5 instead of CAMS, as one the potential improvements in Section 4.4 (lines 741-746).

Also it is unclear whether the PROSPECT derived inputs are different from the existing CLMS biophysical products? And if not, why they are calculated differently?

The biophysical processor (PROSPECT-4SAIL) derives LAI, fg as well as pigments used for leaf albedo estimation. The two first parameters could be obtained from CLMS LAI/fAPAR (biophysical) products but the pigments could not (lines 599-601). While the green LAI derived with biophysical processor agrees quite well with the CLMS LAI (Figure 17, middle panel), the fg derived with both methods are quite different (Figure 17, left panel) with fg from biophysical processor better capturing the expected behavior (Figure 18 and lines 627 – 638). This does not imply that the CLMS LAI is less accurate, just that we needed to use a simplified method to derive fg with the CLMS LAI and FAPAR products. Therefore, since we needed to derive the pigments and fg with biophysical processor we used it also to derive LAI.

Validation: Although the authors use a large number of EC stations for the validation, additional evidence is required for the statement "The CLMS ETa prototype also compared favourably with the global WaPOR ETa maps produced by FAO, which it is meant to replace and other higher-resolution ETa datasets (Section 4.1). The addition of ETa product in the CLMS portfolio should therefore significantly enlarge the CLMS user community" Except for figure 16, the paper does not show how datasets compare for larger areas (spatial patterns).
Thank you for this valid suggestion. For the revised manuscript, we will conduct a more thorough intercomparison of spatial patterns in the WaPOR and CLMSE ET maps.

Model vs framework: To improve clarity, I would advise to distinguish more explicitly between the model (the algorithms) and the framework (the processing system including input selection, gapfilling, and temporal aggregation). The paper would benefit from making this distinction as it helps to understand the design choices. For example, in the sentences *"Preparatory activities required 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 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."* The reference should be to the WAPOR ETLook model instead of the WaPOR framework, as the approach is different from the WaPOR modelling framework with regards on input selection, gapfilling and temporal aggregation.
Thank you for pointing this out. We will review the whole manuscript to make this distinction clearer.

SPECIFIC COMMENTS

Introduction

30 - “Since actual evapotranspiration is a direct proxy of plant water use it can be utilized for consistent irrigation water use monitoring across natural and political boundaries”: Since distinguishing between rainfall and irrigation water use remains a challenge, please clarify this limitation or remove the specific reference to irrigation monitoring.
Yes this is true and we will change “irrigation water use monitoring” to “agricultural water use monitoring”.

50 - “In order to satisfy this wide range of potential users’ needs, and for consistency with other global CLMS products, the CLMS ETa product will have a spatial resolution of 300 m and a dekadal temporal resolution.” It is not entirely clear to which users the 300m product caters?
Different applications which could benefit from ETa dataset are mentioned on lines 32-39 with further applications of WaPOR data mentioned on lines 43-45. When performing analysis on global or (large) regional extent the resolution of 300 m is in many cases sufficient. Those are the user needs referred to in the sentence. However, we will try to modify this statement to avoid confusion. In addition, we will add a paragraph in the introduction explaining the constraints imposed on the global CLMS products, including the 300 m resolution.

55 - “Another operational and global product which utilizes MODIS and VIIRS data is produced by United States Geological Survey using SSEBop energy balance model (Senay et al., 2020) with dekadal temporal resolution and 1 km spatial resolution.” Consider mentioning FEWS as the dataset is available there.
Thank you for pointing this out, we will mention FEWS in the revised manuscript.

Data and methods

Table 2: Perhaps specify which inputs are used for which model? Personally I think a figure showing how these inputs are used to generate the model inputs (e.g. LAI, albedo) would give more insight. I assume “100m” in the weather data means “at 100m above the surface” and not to the spatial resolution – this may be made more clear, or removed.
Thank you for the suggestion. In the revised manuscript we will make it clearer which inputs are used by which model, either by modifying the table, adding a figure or explanatory text. And indeed, 100m refers to 100m above the surface and this will also be clarified.

110 - “same or similar values in both cloudy and sunny conditions (e.g. leaf area index does not change day to day depending on cloudiness). Therefore, gaps in this data are highly suitable for filling using spatio-temporal gap-filling” I understand this makes the data suitable for temporal gap-filing, but it does not automatically make it suitable for spatial gapfilling**?**
The second sentence refers to the full first sentence, the first half of which (“The reflectance

values, as well as biophysical traits, usually show a clear seasonal cycle and spatial similarity “) is on the previous page (line 107) and therefore could have been missed. The spatial similarity makes the data suitable for spatio-temporal gap-filling (such as StarFM which was referred to) which is different to spatial gap-filling (such as interpolation between known data).

121 - Please also introduce View Zenith Angle (VZA) in the text (it is currently only in the captions).
Yes, we will

270 - “More details and the list of evaluated indices are available in the WaPOR wiki (https://bitbucket.org/cioapps/wapor-et-look/wiki/Intermediate_Data_Components/LST, last accessed: 22/07/2025)” : This repository recently moved to https://github.com/un-fao/wapor-et-look , consider updating.
Thank you for pointing this out, we will update accordingly.

275 - “Finally, since we do not expect strong influence of aspect and slope on LST those two variables were removed from the WaPOR list and the resulting combination of 9 variables (called "DMS - WaPOR selected" in Section 4.3.2) is used in the ETa processing chain to sharpen the 1 km Sentinel-3 LST to the required 300 m spatial resolution.” On what did you base this expectation?
From physical considerations, the slope and aspect influence LST mainly through the solar illumination geometry and this is already partially taken into account in the cosine solar inclination angle. We will add this clarification on lines 275-276 in the revised manuscript. In addition, we performed evaluation of LST sharpening using all WaPOR explanatory variables and selected WaPOR explanatory variables (i.e. without slope and aspect) which is presented in Section 4.3.2 and Table 10. Our expectation was based on this analysis.

255 - “ETLook model (Bastiaanssen et al., 2012) is used in the WaPOR framework and is described in detail in Section 5 of "WaPOR Data Manual, Evapotranspiration v2.2" (FRAME Consortium, 2020). “: Please note that the WaPOR data manual refers to the ETLook version 2, and mostly describes how the inputs are derived, while the methodology (the model) used in v2 is described in the methodology document (https://openknowledge.fao.org/server/api/core/bitstreams/d3db4794-fb5b-444c-9b3a-c5fb154c5f9f/content). For version 3 the data manual and methodology documentation were combined, with all updates and changes described in the Github page.
Thank you again, we will also update this accordingly.

3 Prototype product validation

Figure 5: The lack of Eddy Covariance stations outside Northern America, Europe and Australia is an issue, with Africa, Asia and South America only represented by a few stations. This is mentioned in the text, and counterargued with that all major climate zones and plant functional types are represented by at least one EC station. In figure 5 you do show the number of dates available for each PFT and climate zone, but could you add the number of stations as well? I think that would improve our insight in which areas are still underrepresented in EC datasets.
Figure 5 actually shows the number of sites available in each PFT and climate zone. We will increase that number for the revised manuscript as we are evaluating new sites in South America, Africa and Asia. Moreover, the analysis will cover a longer period.

400 - Temporal aggregation smooths errors. Why did you choose to validate the dekadal computations?
It is true that temporal aggregation smooths errors. However, the CLMS ET product has dekadal timestep which is why we validated the data at the same timestep.

440 - "Missing data during the day were computed by linear interpolation if the number of valid timeslots during daytime was at least 50% of the total number of timeslots in that period. Otherwise, the day was discarded." This means you are interpolating both inputs and outputs, so I would mention that this interpolation may smooth variability and can influence error metrics.
The interpolation described in this paragraph does not refer to model input or output. It was applied to a reduced number of eddy covariance data that were only available at sub-daily time step. The procedure is intended to ensure that there is a minimum number of actual *in situ* observations in a day before sub-daily values are aggregated to estimate daily ET. We will try to make this clearer in the revised manuscript.

470 - "Conversely, the models performed less good in the Tropical and Dry regions." I would not attribute the poorer performance to the models themselves as the causes are likely input related. In tropical regions, frequent cloud cover will result in missing remote sensing data inputs, while in dry regions it may be a result of missing short-term ET peaks after rainfall.
We will change this to "model outputs" which captures both the models and their input data.

455-485 – The text describing the figures does not describe the WAPOR outputs while they are in the figures. The comparison with WaPOR is available in the discussion section. But since WaPOR is also based on the ETLook model, but uses other inputs, this would be an excellent opportunity to assess the impact of different inputs (sensors, datasets) and different input timesteps (daily vs decadal).
Since WaPOR ET is not a product of this study, we do not think that it should be analysed in the results section. This could be part of another study in which spatio-temporal evaluation of

different products could be assessed in detail (e.g.CLMS-ET, WaPOR, MOD16, GLEAM, LSA-SAF…). At the same time, we put it in the figures to avoid duplication and to make the analysis performed in the Discussion easier. In the revised manuscript, we will either add a clarification on line 455 that WaPOR shown in Figures 6 – 11 will be discussed further in Section 4.1 or remove WaPOR from those plots. Direct use of WaPOR and CLMS ETLook outputs to assess the different inputs might not be straightforward for the reasons mentioned in Section 4.1 (i.e. reanalysis versus NRT mode and that not all temporarily-static, spatially distributed parameters used in WaPOR are public [note that in the revised manuscript we will used new CLMS ETLook data obtained with the use of the dry bare soil and wet-vegetated surface albedo layers and also other parameterization changes and we expect the results to be much closer to WaPOR maps]) but we will try to expand on this discussion. Finally, CLMS uses daily input timestep and the aggregation to dekadal is performed on the output ET maps, as will be clarified in Section 2.9.

Figure 8: I would add the climate region to the individual plots (instead of A, B, C and D).
We will modify the figure accordingly.

Figure 8/9: I would also add the number of sites used for each figure. Or the number of data points. Now they seem to have the same importance while some are based on more data points.
This is a valid point. We do show the number of sites per climatic zone and PFT in Figure 5 and we also mention it when discussing the results (e.g. lines 473-474). However, in the revised manuscript we will make sure that this is clearer.

Figure 10: The reason for selecting the specific validation sites is not fully explained. If these sites are selected to illustrate the difference between the two models, this should be made explicit. The differences between the models (in particular T) requires further discussion. Moreover, I have some concerns regarding the choice for EBF (evergreen broadleaved forest) and DBF (deciduous broadleaved forest) sites as evapotranspiration modelling of forests is rather complicated for any ET model. For readability I would repeat the abbreviations like EBF more often, in particular in figures like figure 10.
The sites shown in Figure 10 and 11 were chosen precisely because modelling of ET in forest is complicated and for this reason the two models gave quite divergent estimates (line 486-489). We agree however that this should be made clearer and we will address this in the revised manuscript. We will also add another section to the discussion in which we look in more detail at the performances of the two models in different ecosystems and climates. Finally, in the revised manuscript we will use updated parameterization of the two models which is expected to reduce the differences between them in the forests.

Discussion

525 - The discussion on the differences between WaPOR Eta and CLMS ETLook Eta is very thorough.
Thank you

538 - “While both CLMS ETa and WaPOR ETa rely on DMS to improve the spatial resolution of LST, the original LST in CLMS is acquired by SLSTR sensor on board Sentinel-3 satellite with 1 km spatial resolution, while the original LST in WaPOR (version 3) is acquired by the VIIRS sensor on board of Suomi-NPP satellite with 375 m spatial resolution.” => WaPOR L1 does not use DMS as VIIRS LST has a spatial resolution of 375m, and DMS would only introduce errors. DMS is only used for WaPOR L2 and L3.
Thank you for the clarification, we will correct the text in the revised manuscript.

543 - Regarding point 2 (WaPOR being a reanalysis product) I have one remark: WaPOR is produced both NRT and after 6 dekads reprocessed. See also https://github.com/un-fao/wapor-et-look/wiki/Understanding%20the%20WaPOR%20Pipeline#wapor-database
This is an imprecise statement from our side. What we meant is that the WaPOR ET data used by us was from the reanalysis dataset because it came from year 2020. This will be corrected in the revised manuscript.

545- Regarding point 3: the tenacity factor of WaPOR ETLook is 2: See https://github.com/un-fao/wapor-et-look/wiki/Release%20Notes & https://github.com/un-fao/wapor-et-look/wiki/Relative%20Root%20Zone%20Soil%20Moisture.
Thank you for pointing this out. In the revised manuscript we will also use CLMS ETLook data produced using maps of surface albedo of full vegetation and dry bare soil provided by the WaPOR consortium and other WaPOR-ETLook parameterization and we expect the results to be more consistent with WaPOR ET.

581 - “The examples in Figure 16 show as well that the number of missing data in the output maps of ETLook and TSEB-PT is larger than in the WaPOR product. The reason for those gaps are the differences between NRT and reanalysis gap-filling (see Section 4.1) but also the different model inputs and treatments of inland water and snow” This explanation should be expanded to include the differences in gapfilling the inputs or outputs. ,
We will mention the WaPOR gapfilling approach in this discussion.

730 - “This situation should be resolved by the end of the decade when Land Surface Temperature Monitoring (LSTM) mission, with a primary objective of frequent monitoring of field-

scale ETa, will join the Copernicus constellation (Koetz et al., 2018). " Is this approach realistic for an operational global product?
Producing an operational global product at ET 50 m (if that becomes the final spatial resolution of LSTM) might not be realistic but the LST could be resampled to 300 m in which case the need for sharpening of Sentinel-3 LST would be avoided.

Thanks!

Annemarie Klaasse