

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 $KcKs \cdot RET$. 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 E_t value for each day, or did you create decadal $KcKs$ values?

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.

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)
- “The CLMS E_t 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 Sentinel-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.

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

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).

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.

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.

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?

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.

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.

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?

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

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.

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?

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.

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.

400 - Temporal aggregation smooths errors. Why did you choose to validate the dekadal computations?

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.

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.

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 dekadal).

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

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.

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.

Discussion

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

- 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.
- 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>
- 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>.

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. ,

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?

Thanks!

Annemarie Klaasse