Dear Prajwal,

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

I thoroughly enjoyed reading this paper; it is very informative and descriptive. I recommend that the editor consider it for publication, subject to addressing some clarification questions and minor corrections. In my view, the paper is somewhat lengthy because several figures present overlapping information, which could be streamlined to make the main text more concise.

1. Other than the requirements mentioned in line 50, which align with the applications of the WaPOR global ETa dataset, I was wondering why a 10-day interval was selected. Since the revisit time of Sentinel-3 instruments is around 2 days at the equator, a finer temporal resolution seems feasible. While the revisit frequency decreases at higher latitudes, the chances of acquiring clear observations, especially during morning overpasses, are already quite low. Therefore, I'm curious why a shorter revisit interval was not targeted. Or is it because of the number of cloud-free images available, as shown in the text? One of the main advantages of using Sentinel-3 over Sentinel-2 for such applications is its more frequent revisit capability. With a 10-day interval, that benefit seems somewhat limited. This is just a question out of curiosity.

There is a bit of misunderstanding here. The temporal resolution is 10-days but that does not mean that the used Sentinel-3 images are acquired every 10-days. In fact, we use every Sentinel-3 image within the 10-day compositing period to derive clear-sky ET estimates at a daily timescale, then gap-fill those to obtain all-sky daily ET estimates and finally composite (average) those to create a dekadal ET map (lines 375-378). The requirement for dekadal compositing is common among many CLMS global products and is partly based on user-needs, partly on cost-efficiency of storying and distributing global datasets, partly on improved robustness of products when temporal averaging is applied. However, the CLMS ET product will also contain TSEB-PT derived instantaneous heat fluxes provided to users at a daily time-step.

To clarify the above points, we will edit the manuscript as follows.

Rename section 2.9 to "Output gap-filling and temporal compositing"

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."

Lines 754-755 in the conclusion will be modified to:

"The product is designed to have a 300 m spatial resolution, dekadal temporal resolution for the water fluxes (evapotranspiration and its components of evaporation and transpiration) and daily temporal resolution for the heat fluxes (latent and sensible heat fluxes), global extent and to be produced in near-real-time with a 2-day delay (Table 1)."

2. How are PAR and NIR calculated? The input appears to include only shortwave and long wave downward, ssrd and strd, respectively, from the CAMS dataset. Are the direct and diffuse PAR and NIR components derived from using a radiative transfer model, such as 6S, mentioned in the text? I think this was not quite clear in the text.

PAR and NIR instantaneous beam and diffuse shortwave irradiance are estimated thanks to the REST2 model (Gueymard 2008) using interpolated CAMS hourly AOD and TCWV as inputs, and then corrected for topography after Aguilar et al. 2010. This will be clarified in the revised manuscript.

References:
1. Gueymard, C.A. (2008). REST2: High-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation – Validation with a benchmark dataset. Solar Energy 82, 272–285. https://doi.org/10.1016/j.solener.2007.04.008.
2. Aguilar, C., Herrero, J., and Polo, M.J. (2010). Topographic effects on solar radiation distribution in mountainous watersheds and their influence on reference evapotranspiration estimates at watershed scale. Hydrol. Earth Syst. Sci. 14, 2479–2494. https://doi.org/10.5194/hess-14-2479-2010.

3. Is Equation 2.4 correct? I think ε (emissivity) should only apply to the outgoing longwave radiation term.

Reflectivity (thermal albedo) is the inverse of emissivity: refl = 1 – ε. Longwave net radiation consists of incoming minus reflected radiation (similarly to net shortwave radiation) plus the emitted radiation. Therefore, the equation is correct.

4. On what basis was the climate classification into Tropical, Dry, Continental, and Temperate made? While presenting the data this way makes sense, it would be helpful to clarify the method used. For instance, most papers use either an already existing climate classification, like the Koppen-Greger, or some aridity indices-based classification.

The Tropical, Dry, Continental and Temperate classes correspond to the main groups of the Köppen climate classification system. In the nomenclature of the Köppen system these groups get the labels A, B, C and D, respectively. We will mention more explicitly in the revised version of the document that those categories were obtained from the Köppen climate classification system.

5. In Table 6, I am not sure why topographic correction was not applied to the total precipitation. Since topographic correction was applied to other variables, it seems this could have been done for precipitation as well, as precipitation is also quite influenced by topography. Especially when downscaling from the 0.4-degree original CAMS resolution to the 300 m, this step might also help to include some local orographic effects. Just a thought.
Topography certainly has an influence on rainfall, even at local scales (i.e. 300 m). However, correcting for this is complex and as far as we are aware not something which can be done operationally at global extent and daily timescale within reasonable computing costs. At the same time, rainfall is not a core input used by either of the two models but is just used to improve the gap-filling procedure.

6. In Line 441, it is mentioned that the in-situ flux data are filtered based on a "realistic range." Could you clarify how this range is determined? Is it based on the solar constant or some other criterion? Or is it just visual inspection?
The realistic range we refer to in the manuscript is intended to consider the situations in which eddy covariance data providers use fill values to indicate that data is missing. For instance, the use of values like "–999.9" for float data types of "255" for the 8-bit integer data type might be found in datasets to indicate missing data. In the data selection step we removed records with fill data in addition to the use of quality flags. In the new version of the manuscript we will rephrase and replace the expression 'realistic range' to avoid misunderstandings.

7. The stress factors utilized in the ETLook model for the soil and canopy involve topsoil and root-zone soil moisture, respectively (Equations 13 and 13b). While the topsoil moisture was approximated using a trapezoidal construction (LST-fc), how was the root-zone soil moisture determined to calculate the stress factor?
The top-soil and root-zone soil moisture are assumed to have the same value and are estimated using the same trapezoid. We will clarify this by modifying lines 371-372 to the following:
"Then the soil moisture of a pixel (representing both top-soil and root-zone) is estimated using the relative location of LST and fC of that pixel within that trapezoid."

8. Figure 6: The caption could be more descriptive (applies to all the figures, in my view caption should be self-explanatory). Additionally, it is not clear how the graph was generated. Were all observed and modeled ETa values across all sites combined to compute the statistics, or were the statistics first calculated per station and then aggregated across stations? For instance, the standard deviation on the x-axis for the EC towers—does it include data from all EC towers?

Thank you for the suggestion, we will revise all the captions to be more self-explanatory. The Taylor diagram in Figure 6 was generated from the full dataset of observed and modelled ETa pairs across all the validation sites. Therefore, the standard deviation on the x-axis, as well as the other metrics in the Taylor diagram, include data from all the eddy covariance stations considered in the analysis. We will also clarify this in the revised manuscript.

9. In addition to point no. 8 for all Taylor plots, in my opinion, when comparing Taylor plots across different plant-function types (PFTs) and climate types, normalizing the standard deviation with respect to the observation could facilitate the comparison. For example, in Figure 8, the in-situ standard deviations differ across climate types. While this does not change the underlying information, normalizing would likely make it easier for the reader to interpret and compare the results.

Thank you for the suggestion. Normalizing standard deviation might make comparison between PFTs and climates easier but it also removes some information from the plot (the actual value of standard deviation). We will consider this suggestion when revising the manuscript.

10. In my view, there are too many figures presenting overlapping information. While I understand that the author wants to illustrate different aspects in the text from different figures, it affects the paper's brevity. This is just my opinion and can be ignored if you disagree. For example, Figure 7: In particular, the PFT graph does not seem to add significant value to the overall narrative and could be moved to the supplementary material, as much of the information is already conveyed in the tables and the Taylor plots. Figure 12: Since it only shows results for TSEB-PT, it could also be moved to the supplementary material.

Thank you for the suggestions and we agree that the manuscript is a bit on the longer side. When revising, we will carefully consider which parts can be moved to the supplementary material including the two mentioned figures.

11. In TSEB-PT, there is no direct control on evapotranspiration from soil moisture. In that respect, I would have expected ETLook to perform at least as well as TSEB-PT in arid and tropical regions. Could you provide some insight or a hint as to why this is not the case? Probably because stress is only coming from the land surface temperature in the ETLook (already LST being accommodated even in the TSEB-PT), rather than any direct soil moisture observation? Could it be a potential outlook incorporating direct remote sensing-based soil moisture?

Both ETLook and TSEB-PT make use of LST to account for soil moisture. ETLook does it more explicitly, due to the derivation of soil moisture using the trapezoid approach, while in TSEB-PT it is more implicit through calculation of the temperatures and energy fluxes of the vegetation and the soil. In the revised manuscript we will add a new section to the discussion where we analyze the assumptions and limitations of the two models through their performance in different climates and plant functional types. Regarding incorporation of direct (radar-based) soil moisture into evapotranspiration models, this is an active research topic, including by some authors of this manuscript, but is out of scope to include here.

12. In line 537, since the paper does not explore the details of any of the reasons mentioned, I would suggest rephrasing "it is obvious" to "the reasons might be," or alternatively, providing a proper citation.
We agree with the suggestions and will change the manuscript accordingly.

13. Figure 14: In my view, this figure could also be moved to the supplementary material.
We will consider this suggestion when revising the manuscript as mentioned in reply to comment 10.

14. Figure 16: The first two rows correspond to May, and the middle row to July. Since all of the locations are in the Northern Hemisphere, is there a particular reason why the same month was not chosen to illustrate the differences in spatial coverage of the dataset?
There was no particular reason apart from the selected spatial and temporal subsets nicely illustrating the points described on lines 581-588. The figure would look very similar if panel B was from May and since this figure is only for qualitative descriptive purposes, the actual date does not matter too much. Therefore we would prefer to keep it as is.

15. In Figure 18, the green line representing green LAI is not clearly visible. Would it be nicer to plot LAI on the top panel and fg in the bottom panel? This would make the figure easier to understand.
In the top panel of each sub-figure, we plot four parameters: green LAI derived in biophysical processor (green line), green LAI derived with use of CLMS LAI/fAPAR (green circles), LAI derived with biophysical processor (black line) and CLMS LAI (black circles). Many of those parameters are close to each other, especially during some parts of the year, which is why not all datasets are always visible on top. The bottom panel shows fg from biophysical processor (green line) and derived with the use of CLMS LAI/fAPAR (green circles). In the revised manuscript, we will try to make the caption and the figure clearer.

16. Additionally, just out of curiosity, could you clarify why the existing CLMS 300 m LAI product was not directly utilized in the model? The intermediate variable LAI produced in this work is also a product, though unpublished at the end, which seems to me like a potentially redundant effort.
The biophysical processor derives LAI, fg as well as pigments used for albedo estimation. The two first parameters could be obtained from CLMS LAI/fAPAR product 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.

17. Figure 23: Does the shaded area represent the spread of the data?
The shaded area represents the uncertainty range (energy balance closure error) of the data. We will add this information to the caption in the revised manuscript.

18. At last, I think it would be very helpful to include a flowchart connecting all the steps from input to output, highlighting the process from obtaining TOC reflectance and LST to calculating ET, if possible.
This is a good suggestion and we will add such figure to the revised manuscript.

Thank you!

Prajwal Khanal