Dear Editor and Reviewers,

Thank you for your editing and comments. We have revised further our manuscript according to your and the reviewers' comments. Please find some of the specific corrections and responses below. The reviewer comments appear in black and our responses appear in blue.

I. Major changes:

- ✓ We have condensed the discussion as suggested by the reviewers. The following paragraphs or sentences are deleted or revised: the first paragraphs in Section 4, the first and third paragraphs in Section 4.1, the third, fourth and fifth paragraphs in Section 4.3, the last paragraph in Section 4.4.1.
- ✓ We add numbering of the subfigures.
- ✓ The difference among the in-situ observations, water balance estimates and ET products in the fourth paragraph of the introduction section is revised to address the reviewer's concerns.
- ✓ The description of the vegetation condition and climate in Section 2.1 are revised.
- ✓ The land cover classification in Table 1 is updated.
- ✓ The statement in section 3.2.1 is revised to avoid potential conflicts.
- ✓ We have revised the supplementary materials by making some necessary edits and adding a new subfigure on the NDVI map of the Tibetan Plateau to illustrate the vegetation condition .

II. Point-by-point response:

1. Reply to the comments by REVIEWER #1

The manuscripts has undergone some substantial improvements, however there are still a few outstanding issues in my opinion which require revisiting. These include:

Reply: We thank you again for the review and the constructive feedback that helps us to improve our work.

- the use of the entire Synthesis ET dataset. It is very clear that this dataset has some major issues regarding consistency. Despite that you "carefully checked it for several times and we are pretty sure about the existence of the temporal inconstancies." doesnt mean you have to use the entire dataset in your analyses. The <2003 and >2018 data seems to be derived from different combinations of datasets (and overestimating <2003 and underestimating >2018). An example how problematic that is shown in figure 2 where the validation results for the Namco site (data available 2019-2020) is very bad for this data product. Authors are advised to carefully reconsider using the entire dataset. Also since none of the validation analyses cover the period before 2003 (used in WB analyse), including this period in the trend analyses may not add to the manuscript (this applies for all products - and therefore the trend analyses prior to 2003).

Reply: Thank you very much for the suggestion. It is true that SynthesisET has some issues and the validation cannot be done for the period before 2003. We also agree with you that SynthesisET data before 2003 and after 2018 should be removed from the analysis. For the inter-comparison analysis of different products (Section 3.2.1), we use the period (2003~2013) when all the products are

available. So, this result is not affected by the issue of SynthesisET. For the trend analysis, since our results are based on 22products, similar conclusions can be obtained as described in Section 3.2.2 even if we remove SynthesisET. We kept showing the SynthesisET data before 2003 and after 2018 in Figure7 to make its temporal inconsistence (extremely high value before 2003 and low value after 2018) visible.

- I am still not convinced about the discussion section which (my interpretation) tries to explain how the different categorised ET products are performing, however, the there is no quantification of how the different categorised ET products perform (as opposed to individual products), it is therefore difficult to validate the statements the authors make regarding the categorised products. In addition, the authors discuss in detail aspects such as parameterisations, model setup, input data etc, however this is not/can not be derived from the research presented. I would strongly advice to reduce the discussion towards aspects that can be supported by the analyses implemented and compared to other research in the field and not to venture into areas that are not substantiated by your research.

Reply: We agree with you to condense the discussion and focus on the issues supported by the analyses implemented. We revised the manuscript accordingly. It is true that we do not directly quantify how the different categories of ET products perform. The accuracy rank of individual products against EC observations is: PMLV2 > ETMonitor > MOD16-STM (PM-type models) > GLASS > SynthesisET (data-driven model) > SSEBop (energy balance model). The accuracy ranking against ET_{wb} is: ETMonitor > PMLV2 > MOD16-STM (PM-type models) > SSEBop (energy balance model) > GLASS > SynthesisET (data-driven models) when comparing We can clearly see the best 3 high-resolution products are PM-type models (Section 3.1 and Section 4.1).. The information on each product is based on literature as described in section 2.2.3 and it helps explaining the ranking.

- the interpretation of the split of the ET data into the different components is also a bit of a stretch. As these datasets are far from validated, the only conclusion that can be drawn is that these sub-datasets are likely in the ranges provided by the different datasets.

Reply: This ET components have not been fully investigated in previous studies and this limits our understanding of the vaporization process. We agree with you that the evaluation of different ET components was limited in this study due to the scarcity of observations on ET components. We did not state that our analysis documents which data product estimates better the ET components, but we can, for example, draw the conclusion on the basis of a land cover map that soil evaporation accounts for the largest share of ET in the TP and it varies considerably among data products and regions. This actually answers the second proposed question in the introduction, i.e. “which processes play a significant role in determining the total ET”.

---

specific comments:

- since you are evaluating 22 different data products the manuscript is difficult to read as in tables/graphs the order of the products keeps shifting around. Also the addition of the methods in table 3 is appreciated, the link to the discussion is not clear (better to classify the products into the same categories as you use in the discussion)

Reply: Thank you, and sorry for the inconvenience for reading. We checked and revised the order of products in all figures following the order in Table 3. The categories mentioned in the discussion are described in the first paragraph of Section 2.2.3.

- ETMonitor in the discussion is classified under PM methods, however from table 3 it appears only Ew and Ess are obtained using PM, but not the main components (Shuttleworth-Wallace) , also GLEAM uses Priestley-Taylor but this data set is described under PM methods

Reply: Both Priestley-Taylor equation and Shuttleworth-Wallace equation are developed based on the Penman-Monteith equation. So, we grouped all of them as PM-types models. This is explained in Section 2.2.3.

- please update the text to mention that for the EC validation only 7 data products were considered (eg this is not clear in line 382)

Reply: We stated that 7 data products were validated by EC observations in Section 2.3.1. We delete this paragraph including line 382 according to your following suggestion (first paragraph of the discussion reads like a summary, that is not the purpose of a discussion section).

- line 270 (caption) what do you mean with the 'data from all five basins were used together'? does it mean you merged the area for the five basins and compared the WB with the total outflow of all these five basins or you appended the timeseries for all basins to calculate the performance indicators? If the first then how did you deal with the short time series for the Heihe river basin?

Reply: We use the second one, i.e., pooling the ET timeseries for all basins to calculate the performance indicators. We kept the evaluation of the WB separate for each basin. We revised the wording to make it clear.

- line 278, what do you mean with pixel-wise ET?

Reply: We mean each pixel in the images.

- line 279 how do you know the two peaks are related to the non-vegetated /sparsely vegetated area (did you evaluated the ET values per land use classes?)

Reply: We deleted this sentence to avoid potential conflicts.

- line 282 I think you mean figure 4b (when presenting more than one figure with one caption, number them, especially if in the text you are referring to only one of the figures presented)

Reply: Thank you. We revised the references to figures and subfigures.

- line 282-283 "large differences" if you look at the SD/mean the differences are about 10-20% do you really consider this large?

Reply: We intended to show that differences across products in the central and western TP are much larger than in the eastern TP. To avoid any potential misunderstanding, we revise it to ‘differences among different products were larger in the central to western TP than in the eastern TP’.

- line 285 'overestimation in arid regions' you havent explained which basins/regions are located in arid climates, then how do you know there is an overestimation in those areas? Plus overestimation

compared to what? what is the baseline?

Reply: We revised this sentence to 'illustrating the relatively larger uncertainty in arid regions'. The regions/basins located in arid climate are detailed in Section 2.1.

- line 290 average ET 'over different products' dont you mean 'of all products'?

Reply: Yes, we revised it accordingly.

- caption of figure 5 the last sentence is disconnected from the previous sentence. Also not clear how much of the area is missing in the regional products (excl areas outside of China)

Reply: The missing area roughly accounts for 17% of the TP. We revised the sentence to 'It should be noted that some products do not have full spatial coverage, e.g., MOD16, FLUXCOM-RS and FLUXCOM-RS-Meteo only provide ET values for the vegetation-covered regions, and two regional products, i.e. MOD16-STM and PMLV2-Tibet, do not cover the regions outside China, accounting for 17% of the TP roughly.'

- first paragraph of the discussion reads like a summary, that is not the purpose of a discussion section

Reply: Agree. We deleted this paragraph.

- line 390-396 this section is not very relevant, so you used two different approaches, this has been done many times before with comparison with in-situ data and water balance as the most used approaches (see section 4.1 of this paper https://hess.copernicus.org/articles/27/4505/2023/)

Reply: Thank you. We deleted this paragraph.

- the following paragraph only evaluates the products that were validated both with the in-situ data and the WB method, but you are completely ignoring the other ET products. How does this limit the conclusions you are drawing?

Reply: It true that this section is based on the 7 products validated by both methods. Considering these products are the main-stream global ET products with high resolution, we think the conclusion here is well documented.

- rest of the same section: I am failing to understand how you can relate the results of the validation to the actual processes happening

Reply: A critical validation requires that the validated data and ground-truth data represent exactly the same ET processes. This may not be always the case, however. We noticed that depending on the definition of ET (e.g., some include canopy rainfall interception loss while some not) and the algorithms, different ET products may represent different processes. For example, some algorithms constrained Rn to positive values to avoid negative ET, which is different from the eddy covariance observation, which can also capture negative latent heat flux due to condensation and deposition.

- paragraph starting line 420. What about the energy balance closure of the EC towers?

Reply: To reduce its impact, energy closure correction was conducted before validation for EC observed half-hourly latent heat flux data, as descripted in Section 2.2.1. In this study, the overall

energy closure ratio is 0.83 (±0.30). The slope (LE+H against Rn−G based on all the available half-hour data) was less than 1 for all site, ranging from 0.44 to 0.83, with a mean of 0.66. This energy closure ratio lays in the range of previous studies (Wilson et al., 2002; Mauder et al., 2024), who showed that the mean imbalance of FLUXNET observations is in the order of 20%.

Reference:

Wilson K., Goldstein A., Falge E., et al. 2002. Energy balance closure at FLUXNET sites. Agricultural and Forest Meteorology, 113(1–4), 223-243.

Mauder M., Jung M., Stoy P., Nelson J., Wanner L. 2024. Energy balance closure at FLUXNET sites revisited. Agricultural and Forest Meteorology, 358, 110235.

- line 432: did you consider storage changes?

Reply: Yes. As described in Section 2.2, when estimating ET using the water balance, changes in terrestrial water storage were derived from GRACE data.

- line 447 'we found that PM models incorporating soil moisture' is the incorporation of soil moisture the only difference between these models? you can only make such statements if/when you create two datasets using the same method one with and one without soil moisture.

Reply: Thank you. Some models (e.g, ETMonitor and GLEAM) require soil moisture as input, thus we cannot remove soil moisture from the model. We revised the text accordingly. It is revised to 'For instance, to improve the accuracy of ET estimates, ETMonitor, which achieves high accuracy in the TP, uses high-resolution soil moisture data to refine the parameterizations of soil and canopy surface resistances to estimate soil evaporation and plant transpiration'.

- line 490, but arent these products specifically made for climate studies?

Reply: Yes. We deleted this sentence.

- line 535-565 how is this all based on the research presented in this paper?

Reply: We understand your concerns. This section was intended to discuss the reasons that cause the discrepancy in the ET components across different datasets as shown in Section 3.3. Although we do not have direct results to evaluate the impact of these factors, this discussion will inspire our further study to explore the ET components and contribute to better ET partition methods.

---

2. Reply to the comments by REVIEWER #2

The revised manuscript by Zheng et al. is streamlined and conveys the scientific message more clearly than the previous version. The authors compare different ET products regarding their usability in the Tibetan Plateau and find that high-resolution datasets perform better than others, despite significant discrepancies between models. In addition, models incorporating dynamic vegetation cover and water stress modules to predict ET outperform those without these features. Although land surface temperature-based remote

sensing models generally perform worse, they tend to do well in arid regions. Additionally, regional parameterization may be key to improving ET products for specific regional applications. These findings add value to the existing knowledge about the strengths and weaknesses of ET products and their usability in the Tibetan plateau. Nevertheless, the paper is quite lengthy, and I believe it could be made more concise to enhance clarity and readability. I suggest making these revisions before publication and recommend this to the editor.

Reply: We thank you for the positive and constructive feedback that helps us to improve our work.

Please find some of my suggestions here:

1) I believe it is necessary to include a land cover map in the supplementary section (omit if already added) and refer to it when justifying certain results, such as in lines 242-244, 282-286, and 291-293. Currently, land cover is introduced in Table 1, which does not align with the descriptions in the justifications, such as "densely" or "sparsely" vegetated, or "arid."

Reply: A land cover map was included in the supplementary materials, which was extracted from the ESA CCI global land cover map. The ESA CCI land cover classification is also added in Table 1, while the actual land cover according to the field survey is shown within the brackets. We also checked and revised the manuscript to make the description of land cover classes consistent. The multi-year averaged NDVI map was also added in the supplementary materials to have a better illustration on the vegetation condition. The dense vegetation cover (generally with high NDVI) is located mainly in the eastern TP, and sparse vegetation cover or bare land (generally with low NDVI) located in the central and western TP.

---

2) For Section 3.2.2, "Temporal Variability in ET Across the TP," since a basin-wide analysis is provided, I believe more emphasis should be placed on basin-wide discrepancies. The following key points could be addressed in the text, but are not limited to:

a) Please clarify whether the analysis was conducted for every product in each basin or if only the median of the products was analyzed across basins.

Reply: We revised accordingly. We conducted the analysis for every product in each basin. But, it would have been too detailed and possibly confusing to present the analysis in detail for each basin and every product. As regards the monthly variability, we described the seasonal evolution of ET in different basins (Fig.6). As regards the yearly variability, we only focus on the statistics for the entire TP and we provided the results for each basin in the supplementary materials.

---

b) Are there any variations in the profiles between the basins? If so, to what extent?

Reply: Yes. As stated in the second paragraph of section 3.2.2: “At the basin scale, the difference in annual trends between different products is also clearly illustrated (Supplementary Figure S6). Most basins showed a significant increasing trend of ET, especially in the Yellow, Yangtze, Mekong, Tarim, Hexi Corridor, Tarim, and Qaidam basins, where most products had a positive ET trend. The median ET trend is either negative or close to zero in the Ganges, Brahmaputra, Amu Darya, and Inner TP basins, probably indicating a decreasing or non-monotonic trend for these basins.”.

---

c) How do the basins differ from each other in terms of land cover or other characteristics?

Reply: The characteristics of each basin are described in Section 2.1. The five basins (Hexi, Tarim, Qaidam, Amu Darya, and Inner TP) located in the northern, western, and central parts of the TP receive relatively low precipitation with arid or semi-arid climate. These basins are generally covered by sparse vegetation or bare land (Supplementary Figure S1). The remaining basins receive high precipitation, and they are characterized by relatively dense vegetation

---

d) Do all basins, regardless of land use, exhibit the same level of uncertainty?

Reply: Surely no. The uncertainty of each basin is detailed in Supplementary Table S2. As stated in Section 3.2.1, "The uncertainty on ET is highest in basins with low ET and sparse vegetation cover, i.e. the Qaidam, Inner TP, Hexi Corridor, Tarim, and Amu Darya basins. The uncertainty is expressed as the ratio of standard deviation to the mean values (Supplementary Table S2). Uncertainty is also high in the Indus and Brahmaputra basins, most likely due to their complex topography, extreme altitude range, and large areas of permanent glaciers and snow, which make it difficult to obtain reliable estimates.".

---

3) Line 76-79: In my understanding, basin-scale ET estimates represent the net water vapor flux from land to atmosphere (whether positive or negative), as they are derived from the water balance. Hence, they differ from in situ measurements, which capture both upward and downward water vapor flux. So, sentence needs restructuring.

Reply: We revise it to 'These evaluations have generally been based on either in-situ measurements using eddy covariance systems or basin-scale ET estimates using water balance method. The *in-situ* eddy covariance measurements at 30min temporal resolution capture both upward and downward water vapor flux at site scale, and the integrated daily or monthly ET depended on whether the upward water vapor flux is included. Water balance estimates capture the net liquid water flux at the surface and at basin scales, while ET products are estimates of the upward water vapour flux, unless separate estimates of condensation and deposition are provided. This difference contributes to the uncertainty.'.

---

4) Lines 85-89 and 90-94 convey the same meaning and can be merged for conciseness.

Reply: Thank you. Revised accordingly.

---

5) Figure 2: Please indicate overlap meaning consistent time frame, 2003-2013, it it's the case.

Reply: Revised accordingly.

---

6) Line 279-280: Cross reference missing.

Reply: Revised accordingly.

---

7) Figure 9: These differences, at least in sublimation, could be due to how snow extent is treated in each model. For instance, it seems that GLEAM considers snow extent differently than ETM, or could this simply be an effect of plotting?

Reply: It is true that the difference of snow/ice sublimation can be partly explained by differences in the input snow cover extent or fraction. ETMonitor use MOD10 snow/ice cover data at 500m resolution as the boundary conditions for sublimation estimation at global scale. If a pixel is covered by snow/ice at certain day according to MOD10 date, sublimation is estimated by ETMonitor. GLEAM use GLOBSNOW daily snow water equivalent at 25km resolution for the Northern Hemisphere and NSIDC monthly snow water equivalent climatology product at 25km resolution for the Southern Hemisphere.

---

8) Section 4.1 Title The paragraph does not align with the content, as it does not address which vaporization processes are more relevant. Instead, it focuses on the differences between processes captured in in-situ and basin-wide data, and how these differences impact ET comparisons.

Reply: We revised the title to 'Contribution of the study to a better understanding of the vaporization processes.

---

9) Line 417-421: I find this quite interesting. Nice finding.

Reply: We thank you for positive feedback.

---

10) Line 437-439: Is not PM also an energy balance model?

Reply: It is true that PM equation considers the surface energy balance since PM equation fundamentally combine the energy balance and the mass transfer principles. The wording "surface energy balance model" generally refers to algorithms like SEBS or SSEBop that utilize land surface temperature to estimate sensible and latent heat fluxes.

---

11) Section 4.2.3 Title Since this section does not only incorporate ensemble ET products, please consider changing title here.

Reply: We deleted 'ensemble' in the title.

---

12) Section 4.3: Page 26: I believe this section could be condensed. While I enjoy reading this, it doesn't seem to align with the focus of the study and could be made concise without writing in such details. Alternatively, for each model used, a supplementary table could be created to detail differences in forcing, model structure, calibration, spatial heterogeneity, and other relevant factors and refer it.

Reply: Thank you for positive feedback. The section is meant to provide information useful to understand some of the observed differences in the accuracy and ET partitioning of data products. In any case we revised this section to condense it.

---

13) Line 582- 588: I believe this resembles the previous version, as the information comes across abruptly to the reader. Since the related section has been removed in the current version, it now lacks the justification behind the statement.

Reply: We deleted these sentences.

---

3. Reply to the comments by REVIEWER #3

Reply: We thank you again for the review and the constructive feedback that helps us to improve our work.

‘2020), Recent studies’, here should be "." not ","

Reply: Revised accordingly.

"certainly contributes to the understand of the ET process" should be "certainly contributes to the understanding of the ET process"

Reply: Revised accordingly.

"Calibration of model parameter" should be "Calibration of model parameters."

Reply: Revised accordingly.

We would like to thank the reviewer for their constructive comments on our work that enriched our manuscript.

With kind regards,
Chaolei Zheng, on behalf of co-authors