This manuscript evaluates 5 different uncertainties estimation techniques using both synthetic tests and flux-tower observations. It shows that EIVD outperforms other uncertainty quantification method. Based on the uncertainties of different ET products, a newly merged ET product is proposed.

Overall, I think this is a very interesting paper with solid materials and is a significant contribution to the field of ET uncertainty and merging studies. I would recommend acceptation after considering my comments:

Response:

We sincerely thank the reviewer for your suggestions, and we have made the following changes to the manuscript according to your suggestions: (1) More detailed explanation of the merging method; (2) The performance of merged products, the inputs and result from the simple average (SA) method are comprehensively compared; (3) Modify the corresponding content of the article according to your specific suggestions; (4) The language of the article has been polished to make it more in line with the characterization of scientific papers.

(Reviewer's comments are marked in red, and our responses are marked in black)

(1) Line 30: ET error is larger at 0.25 spatial resolution. Could it be the representativeness error of flux towers?

Response:

We sincerely thank the reviewer for your comment. We added relative analysis in Section 5:

"…Moreover, a slight overestimation of the merged product was found over 0.25° for all land cover types. Since site-based observation could only cover the variation of geographical variables over a certain range (Tang et al., 2018), pixel-based evaluations at 0.1° and 0.25° should consider the increasing representative error of in situ observations, which may explain the increased bias found in our comparison…"

(2) Line 50: "lots of" consider changing it into a more appropriate word in scientific writing.

Response:

We sincerely thank the reviewer for your suggestion. "lots of" is replaced with "many".

Modifications have been made in the corresponding places

(3) Lines 82, 83 and elsewhere: revise the format of the citations.

Response:

Thanks for your comment, we have adjusted the format of the citation

(4) Line 87: “double” => “double instrumental variable algorithm”

Response:

We are truly grateful for the reviewer’s comment. Modification has been made.

(5) Lines 250 to 264: a significant portion of this paragraph should be placed in the introduction.

Response:

We sincerely thank the reviewer for pointing out our mistake. In the new manuscript, we use TC as an example to illustrate the principle of the collocation method and describe the differences between different methods (Section 3.1).

(6) Line 283 and elsewhere: please enumerate the equations

Response:

Sincerely thanks to the reviewer’s comment. All equations have been enumerated

(7) Section 3.2: The merging is aimed to address random errors. The biases should be explicitly handled. However, this is not clear in the current manuscript.

(8) Line 461: There are several products clearly violates the assumptions of TC/QC/IV. For instance, ERA5, GLEAM and GLDAS are all model based. These products should not be used together. Therefore, why Table 4 reports all-method-averaged metrics, instead of only the metrics from “reasonable” product combinations.

Response:

We truly thank the reviewer for your comments, the reviewer's two comments are related, so we reply to them together. In the previous manuscript, there was an error in the description of the merging formula, and we corrected it accordingly (Section 3.2).

“…Given specific variances of inputs, linear combination could serve as a simple and efficient solution for data assimilation. In this study, each product (i) is assigned the optimal weight (ω_i) that minimizes the mean square error (Bates and Granger, 1969;

Kim et al., 2021) using error variances ($\sigma_{\varepsilon_i}^2$) and the ECC ($\sigma_{\varepsilon_i\varepsilon_j}$) as:

$$\omega_{ij} = \frac{\sigma_{\varepsilon_i}^2 - \sigma_{\varepsilon_i\varepsilon_j}\sigma_{\varepsilon_i}\sigma_{\varepsilon_j}}{\sigma_{\varepsilon_i}^2 + \sigma_{\varepsilon_j}^2 - 2\sigma_{\varepsilon_i\varepsilon_j}\sigma_{\varepsilon_i}\sigma_{\varepsilon_j}} \quad (9)$$

The combined product θ_c is calculated as:

$$\theta_c = \sum_{i=1}^{N} \theta_i\omega_i \quad (10)$$

where ω_i is the weighted arithmetic mean for each product. For a dual-input combination, the value of ω_i is calculated as:

$$\omega_i = \frac{\omega_{ij}}{\omega_{ij} + \omega_{ji}} \quad (11)$$

For a triple-input combination, the value of ω_i is given as:

$$\omega_i = \frac{\omega_{ij} + \omega_{ik}}{(\omega_{ij} + \omega_{ik}) + (\omega_{ji} + \omega_{jk}) + (\omega_{ki} + \omega_{kj})} \quad (12)$$

…"

This study uses the forecast combination suggested by Bates and Granger (1969), which suggests using empirical weights based on forecast variances. With further information on the error correlation, this method can work well in practice. We follow the same equation used in (Kim et al., 2021). Since the input ET products are not completely independent, the impact of ECC on the results was fully considered in the study, and the analysis and description of the ECC results was added to the new manuscript (Section 4.3). The selection of subsequent fusion products is also based on product errors and ECC calculation results (Section 4.4). In addition, we have added an analysis of the error of the fusion product to the discussion (Section 6). In the product evaluation part, in addition to the KGE coefficient, we also calculated the RMSE and R^2 coefficients, and randomly selected 4 sites (2 new sites) to compare the performance of the merged product, the input product and the weightless average (Section 5).

(9) Section 4.12 and 4.2: I think they are better suited in the results section.

Response:

We are very grateful to the reviewer for your comment. The purpose of this section is to assess the reliability of the collocation methods by comparing the product error calculated by the collocation methods with the product error calculated on site basis. We have adjusted Section 4 and more proper description has been added accordingly.

(10) Section 5.1: which method is used here?

Response:

We sincerely thank the reviewer for your comment. The description of inputs and used method is included in Section 4.4:

"…The selection of the algorithm was based on the validation results. The data were produced over 0.1°-8Daily and 0.25°-Daily resolutions. Among the methods, IVD, EIVD, and QC were the three preferred methods for usage. Since the QC method requires four sets of data inputs, the requirements cannot be met in the corresponding period. Therefore, we use the error calculation results of the EIVD and IVD methods as the basis for weight calculation. Product selection for different time periods is based on the availability of the product. In this study, each product is assigned with the optimal weight (ω_i) that minimizes the mean square error (Bates and Granger, 1969; Kim et al., 2021) using collocation-evaluated error variances (σ_{ε_i}) and the ECC ($\sigma_{\varepsilon_i \varepsilon_j}$). The table below showed the data used for merging during different periods.

TABLE 2. Combination of inputs and accessible methods

Scenario 1 (0.1°-8Daily)		
Time Period	**Available Products**	**Method**
(2001.01.01-2002.07.03)	ERA5/FLUXCOM	IVD
(2002.07.04-2013.12.27)	ERA5/FLUXCOM/PMLV2	EIVD
(2013.12.28-2019.08.29)	ERA5/PMLV2	IVD
Scenario 2 (0.25°-Daily)		
Time Period	**Available Products**	**Method**
(1981.01.01-2003.02.01)	ERA5/GLEAM	IVD
(2003.02.02-2018.12.31)	ERA5/FLUXCOM/PMLV2	EIVD
(2019.01.01-2020.08.31)	ERA5/GLDAS	IVD

The results of numerical analysis show that the influence of ECC should be fully considered in the process of calculating weights. Since the IVD method cannot be used to calculate ECC, during the period using the IVD method, we use the ECC result obtained by the EIVD method to calculate the weights…"

(11) Figure 11: as commented above, the biases of different products should be removed first, before merging. It may not affect correlations but will have some impacts on RMSE.

(12) Figure 15: likewise, the merging is aimed to reduce random errors and is not expected to improve trends. Theoretically, we should remove all the systematic differences of the parent products prior to merging.

Response:

We sincerely thank for your comment, the reviewer's two comments are related, so we reply to them together. The global trend analysis of ET is not the focus of this article, so this section has been removed in the new manuscript. The new Section 6 focus on the discussion of merged products.