Dear Referee#2,

We greatly appreciate your valuable comments, which have helped us improve our manuscript extensively. We have replied to your comments individually and revised the manuscript accordingly.

For your convenience, please see attached zipped file containing:

- Our responses (your comments are in black while our responses are in blue): *2025-4522_ResponsestoRC2.pdf*
- Revised manuscript with tracked changes : *2025-4522_RevisedManuscript_withTC.pdf*
- Revised supplementary document with tracked changes : *2025-4522_RevisedSupplement_withTC.pdf*

Thank you very much.

Sincerely,

Mwangi et al.

Referee Comments (and our responses):

This manuscript presents a comprehensive analysis of uncertainties in long-term ensemble estimates of contextual evapotranspiration (ET) over Southern France using the EVASPA modelling framework and MODIS remote sensing products. The study addresses an important and timely topic, as uncertainty quantification in satellite-based ET estimates remains a major challenge for hydrological and climate applications.

The ensemble-based approach adopted by the authors is methodologically sound and well aligned with current best practices in remote sensing of land–atmosphere fluxes. The manuscript demonstrates a strong technical foundation, and the comparison with flux tower observations adds credibility to the analysis.

Thank you very much for taking the time to read our work and providing valuable feedback. We have addressed your concerns accordingly; please see our comments below and the modifications within the revised manuscript.

Major comments

- A key issue is the comparison against the daily ET estimates measured by flux towers. Is their footprint being compared with the pixel dimension? A scale issue should always be addressed when discussing such comparisons.

Thank you for your comment. We do agree that there is a likely mismatch between the tower footprint and the pixel size. We have added text acknowledging the scale issue introduced when ET estimates are compared against in-situ EC data.

L696-704:

"Whereas this study is generally appropriate for site-based uncertainty assessments, we also acknowledge that several spatial scale issues exist that limit spatial generalization and transferability of the results. In particular, there is a clear scale mismatch between the estimates derived from (and for) the MODIS pixel, which represent relatively large and heterogeneous spatial areas, and the in-situ observations, which are representative only of the localized EC flux tower footprint. As such, the performance assessment results should be interpreted with caution, as the comparison involves estimates and observations with fundamentally different spatial coverage and representativeness. This limitation is further compounded by the presence of mixed pixels, where multiple land cover types and surface characteristics are aggregated within a single pixel, introducing additional uncertainty that is not explicitly accounted for in our analyses."

- Furthermore, even though the mail goal here would be the spread of the "error", more than the error itself, it would be helpful to sort some results in terms of land cover, as you did with the four seasons, in order to assess potential for uncertainty in different environments.

In addition to seasonal analyses, we have added results comparing results in terms of land cover. Thank you for the suggestion.

L560-563:

"The heat maps corresponding to the different land covers and seasons are presented in the supplementary document. Across the land covers, the trends are generally similar to those in Figure 9Figure 98Figure 8 with only subtle differences (Figure S 4, supplements)."

L618-626:

"Only the bar graphs for all sites combined are shown - the site-specific graphs show a similar trend (see Figure A 2). From the plots, the overall trends (particularly on ranking of the four variables) appeared largely similar across the analyses, even when alternative reference combinations of the distinct variables were considered. This suggests that the observed patterns are largely consistent and not overly sensitive to the specific choice of reference inputs-methods combination. LST still appears to introduce most of the overall (all seasons) uncertainty over the three land covers (see supplements, Figures S 6, S 7 and S 8). Nonetheless, the salience of EF methods on the uncertainty over the grasslands and croplands is apparent, especially in the non-interpolated daily ET estimates during summer."

Supplements:

In addition to Figures S 6, S 7, S 8 recalled below, also see Figure S 4, where the heatmaps are separately generated according to land cover type.

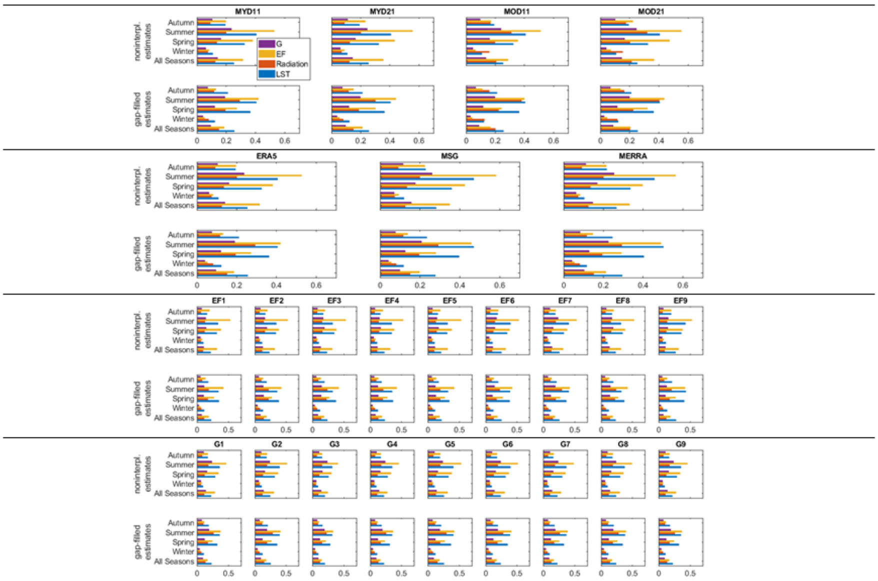

Figure S 6: As in Figure 12, but for Croplands – Auradé and Lamasquère.

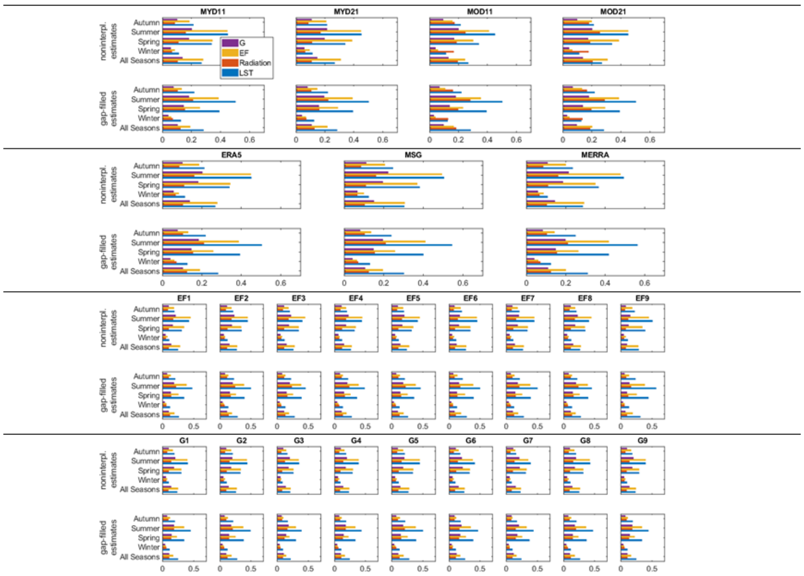

Figure S 7: As in Figure 12, but for Forests – Bilos, Le Bray, Fontblanche, Puechabon.

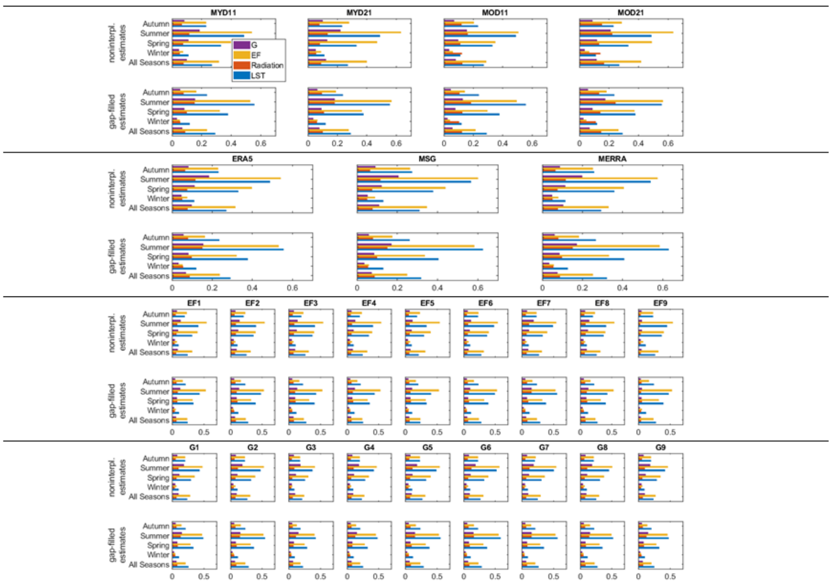

Figure S 8: As in Figure 12, but for Grasslands: Coussoul, Toulouse.

- L185-186. This assumption is quite common throughout the literature. However, it would be helpful to recall somewhere what uncertainties lie behind it, and, if possible, try to include them in the uncertainty quantification you obtain later from your ensembles

  Thank you for the clarification. Indeed, it is a common assumption that λe-to-solar radiation ratio is conserved throughout the day, which introduces further uncertainties in the estimated daily evapotranspiration estimates. We have updated and clarified the text accordingly to acknowledge such limitations.

  L203-216:

  "Eventually, the daily ET_d $(\lambda E/\lambda)$ $[mm/d]$ is computed using the daily solar radiation data, where it is commonly assumed that the λE-to-solar radiation ratio is conserved throughout the day, i.e. $\frac{\lambda E_I}{SW_I} \approx \frac{\lambda E_d}{SW_d}$ (e.g., Delogu et al., 2012)."

  "It should be noted this is one of several methods used to aggregate latent heat fluxes to the daily scale. It was selected herein as it is physically based, requires minimal data and relatively simple to apply. We nonetheless acknowledge that it does introduce uncertainties in the estimations, especially since the assumption that the ratio is conserved throughout the day may not always hold. Within the TRISHNA SWEAT framework, other methods (e.g. conservation of the evaporative fraction throughout the day) are being tested to quantify such uncertainties between methods, and for possible inclusion as part of the product chain."

- Since the focus here is uncertainty, I would have liked more quantitative references to the standard deviations of the different combinations (e.g., in Fig.5 it seems a missing information), both in absolute and relative (to the main variable values) terms. Furthermore, as seen in Fig.6, a lot of the ensemble results seems to distribute quasi-normally, although a non-negligible quantity (e.g., Aurade, Caoussoul, LeBray, Toulouse) do not. Given these distributions, a skewness indicator or other, more complete parameters would seem to be better fit to describe the results of the different indexes you computed, as opposed to the simple SD, which can be misleading when the sample is not distributed normally.

  Well noted. We acknowledge that standard deviation (SD) can be misleading/misinterpreted when used with non-Gaussian data. We however preferred using standard deviation throughout the main text since SD will be used as the primary uncertainty measure in the ensemble estimates within TRISHNA. Note that the interquartile range (IQR), which is useful in quantifying non-normal/non-symmetrical spreads, is already qualitatively depicted in the box plots displayed within the violin plots, Figure 6. We have now added the quartile coefficient of dispersion (and the coefficient of variation) to complement the SD.

  L235-240:

  "To be consistent with recent ensemble ET products, which quantify uncertainty in SD, the TRISHNA ET products will also include the ensemble SD as the primary uncertainty measure (this may change as the products evolve). We nonetheless acknowledge that SD (and the corresponding coefficient of variation, CV) can be misleading, especially when the analysed data exhibits non-symmetrical (specifically non-Gaussian) distribution. As such, the coefficient of quartile variation or quartile coefficient of dispersion (QCD)—which was recommended for non-normal distributions by Bonett (2006)—has also been included as a complementary statistical measure of uncertainty in the EVASPA daily ET estimates.”

L315-319:

“

$$SD_i = \left(\sum_{j=1}^{n} \left(Y_{i,j} - \mu_i \right)^2 \Big/ n \right)^{0.5} \; ; CV_i = SD_i / \mu_i \qquad (5.iii)$$

$$QCD_i = \left(Y_{Q3,i} - Y_{Q1,i} \right) / \left(Y_{Q3,i} + Y_{Q1,i} \right) \qquad (5.iv)$$

”

“... μ and n in the SD are the ensemble mean and ensemble size, respectively. $Y_{Q1,i}$ and $Y_{Q3,i}$ in the QCD are the respective first and third quartiles at i^{th} time step”

L442-451:

“The relative uncertainty (see CV and QCD) within the ensemble is also highest at this site. Conversely, the model also seems to overestimate the ET fluxes at the other grassland site (Coussoul). Although this site is relatively homogeneous in land cover (and land use), it is located in close vicinity to some water ponds, which may interfere with the thermal signals captured at the pixel scale. This likely introduces noise that propagates through the modelling process influencing the final estimates. As expected, seasonal performance of daily ET across all EVASPA member sets is consistently driven by the availability of radiation used for the surface turbulent fluxes, with the best performance (in RMSD and MAE) observed in winter. The performance then cyclically degrades over spring through summer and improves into autumn (Table 2, Figure 6). The absolute uncertainty (SD) in the estimated ensemble ETs shows a similar cyclical trend. The more informative relative uncertainty (as given by the relative SD or CV and the corresponding QCD) is however generally consistent throughout the seasons.”

Table 2: Global performance of estimated daily evapotranspiration (average of all EVASPA ensemble members). Uncertainties within the ensemble are quantified using the median values of the standard deviation (SD), coefficient of variation (CV) and quartile coefficient of dispersion (QCD).

	Interpolated daily ET estimates							non-interpolated daily ET estimates						
					*uncertainty (median)**							*uncertainty (median)**		
	RMSD [mm/d]	*Willmott's D [-]*	*Bias [mm/d]*	*MAE [mm/d]*	*SD [mm/d]*	*CV [-]*	*QCD [-]*	*RMSD [mm/d]*	*Willmott's D [-]*	*Bias [mm/d]*	*MAE [mm/d]*	*SD [mm/d]*	*CV [-]*	*QCD [-]*
Aurade	0.93	0.83	0.09	0.68	0.40	0.30	0.21	1.19	0.80	0.22	0.90	0.47	0.24	0.16
Bilos	0.84	0.86	-0.17	0.62	0.41	0.33	0.24	1.10	0.79	-0.08	0.84	0.50	0.28	0.20
Coussoul	0.97	0.63	0.58	0.72	0.46	0.40	0.27	1.15	0.58	0.71	0.90	0.53	0.37	0.26
Fontblanche	0.93	0.58	-0.19	0.74	0.38	0.40	0.27	1.07	0.52	-0.01	0.85	0.42	0.36	0.24
Lamasquere	1.04	0.70	-0.41	0.70	0.40	0.43	0.30	1.27	0.62	-0.37	0.89	0.48	0.37	0.25
LeBray	0.74	0.85	-0.11	0.56	0.41	0.35	0.24	0.88	0.83	0.06	0.68	0.47	0.29	0.20
Puechabon	0.87	0.84	0.44	0.67	0.42	0.25	0.17	1.09	0.81	0.57	0.85	0.47	0.22	0.15
Toulouse	1.14	0.53	-0.75	0.82	0.38	0.69	0.48	1.35	0.38	-0.83	0.99	0.44	0.66	0.45
Croplands	0.99	0.78	-0.15	0.69	0.40	0.36	0.25	1.23	0.73	-0.07	0.89	0.47	0.29	0.20
Forests	0.87	0.80	-0.01	0.67	0.40	0.33	0.23	1.06	0.76	0.17	0.83	0.46	0.27	0.19
Grasslands	1.09	0.49	-0.32	0.79	0.41	0.51	0.35	1.27	0.40	-0.21	0.95	0.49	0.45	0.30
Winter	0.52	0.54	-0.14	0.39	0.16	0.35	0.25	0.60	0.49	-0.07	0.47	0.18	0.30	0.21
Spring	1.02	0.71	-0.23	0.79	0.52	0.35	0.24	1.25	0.63	-0.13	0.97	0.57	0.31	0.21
Summer	1.32	0.60	0.04	1.04	0.70	0.32	0.22	1.52	0.57	0.20	1.22	0.70	0.32	0.22
Autumn	0.70	0.66	-0.05	0.54	0.30	0.34	0.24	0.83	0.60	0.07	0.65	0.34	0.31	0.22
All sites and seasons combined	0.94	0.77	-0.09	0.69	0.40	0.34	0.23	1.16	0.73	0.03	0.87	0.47	0.31	0.22

** Reported uncertainty are the median values over the uncertainty time series';*
E.g. for SD, the standard deviation is calculated for the ensemble at each time step, then the median of the SD time series is reported.

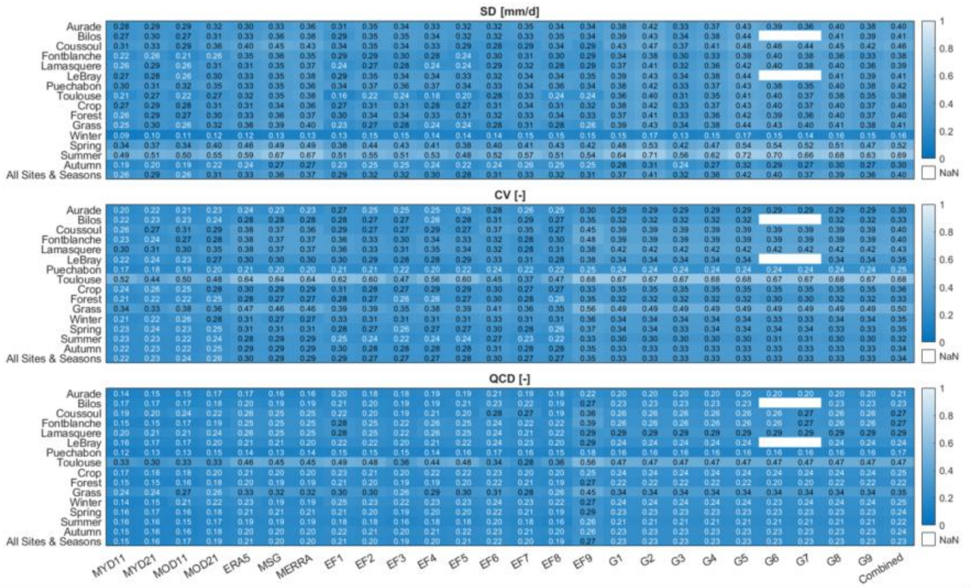

Figure 6: Uncertainty heat maps (median of SD, CV, and QCD over the study period) for the different EVASPA models across all evaluation sites. More details on how the sets are constructed are contained in Figure S 1 in the supplements.

See supplementary document for results including VIIRS based estimates

Minor comments

- No need to replicate the LST and NDVI acronyms, you already mentioned them at L54

  We have removed the unnecessary/repeated acronyms. Thank you.

- This "relatively low" maybe would be better supported by some mention of relative uncertainty (standard deviation / average), to help the reader in contextualizing the potential error

  Supporting text quantifying the (relative) uncertainty has been added. We have added relative measures of uncertainty in the text, i.e. the coefficient of variation CV = standard deviation / average, and the coefficient of quartile variation (also quartile coefficient of dispersion, QCD)

  L236-240:

  "We nonetheless acknowledge that SD (and the corresponding coefficient of variation, CV) can be misleading, especially when the analysed data exhibits non-symmetrical (specifically non-Gaussian) distribution. As such, the coefficient of quartile variation or quartile coefficient of dispersion (QCD)—which was recommended for non-normal distributions by Bonett (2006)—has also been included as a complementary statistical measure of uncertainty in the EVASPA daily ET estimates."

  L315-319: "

$$SD_i = \left(\sum_{j=1}^{n} \left(Y_{i,j} - \mu_i \right)^2 / n \right)^{0.5} \; ; CV_i = SD_i / \mu_i \qquad (5.iii)$$

$$QCD_i = \left(Y_{Q3,i} - Y_{Q1,i} \right) / \left(Y_{Q3,i} + Y_{Q1,i} \right) \qquad (5.iv)$$

”

“… μ and n in the SD are the ensemble mean and ensemble size, respectively. $Y_{Q1,i}$ and $Y_{Q3,i}$ in the QCD are the respective first and third quartiles at i[th] time step”

- On a more general note, the manuscript is at times a bit too dense. Although this is understandable, given the scope of the analysis, this can hamper readability. Some sections, especially Section 4.2, feature long paragraphs with many interlinked concepts. Also some figures (e.g., 8, 10 or 11) feature either a complex caption, too many information or are hard to interpret.

  Thank you for the feedback. The long sentences have now been shortened to improve readability. The captions have also been rephrased to make them clearer. Please refer to the revised manuscript with ‘tracked changes’.

  For example,

“

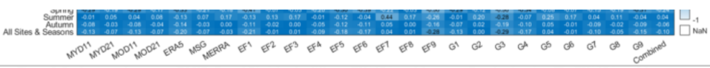

Figure 5: Performance heat maps (RMSD, D, and Bias) for the different EVASPA models across all evaluation sites. The EVASPA models are composed of all gap-filled member estimates in the set (e.g. the MYD11 set is constructed using the MYD11 LST/E data together with the 3 radiation datasets, 9 EF and 9 G methods). More details on how the sets are constructed are contained in Figure S 1 in the supplements.

Figure 8: Heat maps of Willmott’s D index *[-]* and MAE *[W.m⁻²] (or MAE [-] for EF)* between the EVASPA SEB members (net radiation - Rn, ground heat flux - G, latent heat flux – LE, and evaporative fraction - EF) over all sites and seasons combined. Heat maps of the metrics are calculated by taking each individual member vs all other ensemble members. *Numbering in the heatmaps/ensemble is done according to the list of variables given in Table A 1. **Distinct number of variable combinations in the heat maps are given in Table A 1.

** E.g. for EF: 1-9 – all 9 EF methods using MYD11 LST, 10-18 – all EF methods using MYD21 LST, 19-27 – all EFs using MOD11 LST, and 28-36 – all EFs using MOD21 LST*
*** I.e., net radiation - 3 × 4 combinations [radiation × LST products] ; G flux – 9 × 3 × 4 [G methods × radiation × LST products] ; LE fluxes involve 9 × 9 × 3 × 4 combinations [G methods × EF methods × radiation × LST products] ; EFs - 9 × 4 combinations [EF methods × LST products]*

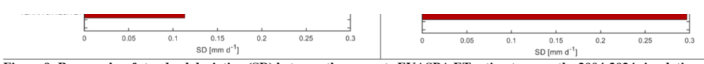

Figure 9: Bar graphs of standard deviation (SD) between the separate EVASPA ET estimates over the 2004-2024 simulation period, i.e., the specific uncertainty for all four distinct variables (across all sites combined). *a)* At AQUA overpass – without and with gap filling / temporal interpolation. *b)* At TERRA overpass – without and with gap filling / temporal interpolation. *c)* For the 4 LSTs (separately between 11 and 21 for TERRA, between 11 and 21 for AQUA, between AQUA and TERRA for 11, between AQUA and TERRA for 21) – c.*i)* based on non-interpolated estimates, c.*ii)* based on interpolated estimates.

”

References

Bonett, D. G. (2006). *Confidence interval for a coefficient of quartile variation*. *50*, 2953–2957. https://doi.org/10.1016/j.csda.2005.05.007

Delogu, E., Boulet, G., Olioso, A., Coudert, B., Chirouze, J., Ceschia, E., … Lagouarde, J. P. (2012). Reconstruction of temporal variations of evapotranspiration using instantaneous estimates at the time of satellite overpass. *Hydrology and Earth System Sciences*, *16*(8), 2995–3010. https://doi.org/10.5194/hess-16-2995-2012