We thank the reviewers for their time and suggestions to improve the manuscript. Our response can be found below the comments in blue.

On behalf of all authors,

J. R. Thomson

Reviewer 1

A well-written and useful study, and a clear contribution that fits ESSD well. The harmonized 14-dataset dataset, the trend intercomparison, and the trend-signature topology framework are genuinely valuable, and the central message, that ET trend estimates are strongly dataset-dependent and that the disagreement is structured rather than random, is well supported. I have little to add: my comments are all minor and none affects the validity of the analysis.

Response: We really appreciate your positive feedback. We are pleased you find the study useful!

Comments

1. The role of FLUXNET (l. 212-213, l. 452-454 and l. 472-474). These passages rest on FLUXNET influencing the datasets more than it does, and I think the statements should be revised.

At l. 212-213 regions lacking flux towers are said to be "likely weakly constrained by direct measurements", and at l. 452-454 the disagreement in "data-rich regions" is called striking. Both assume that flux-tower density constrains the datasets, but in-situ ET is not an input to any of them, so it is not clear that it does. Please clarify what "data-rich" means here, and state the mechanism by which tower density would constrain the ensemble, or revise the claim.

At l. 472-474, shared FLUXNET use is said to make datasets converge. Convergence would follow from shared calibration, not from shared evaluation, and the "or evaluation" makes the claim almost always true. The four datasets named are process-based or semi-empirical and, as far as I know, use FLUXNET mainly for evaluation, so the convergence argument does not hold for them.

Response: Thank you for your comments. We address them jointly below.

Flux tower networks such as FLUXNET are not only used for evaluation but also for calibration. Even in process-based or semi-empirical models, there are parameters that need to be tuned or calibrated. Additionally, machine learning modules used within ET datasets are often built around and trained with flux tower information. The following is a more detailed description of how BESS, ETMonitor, GLEAM and MODIS utilize flux tower stations for calibration. BESS V2 uses FLUXNET to calculate plant-functional-type-related parameters in its terrestrial ecosystem respiration (TER) module (Li et al., 2023), but the authors also note: “We used the entire recently released FLUXNET2015 Tier 1 dataset, covering a total of 206 flux sites and 1496 site-years (Pastorello et al., 2020), for model calibration and validation.” ETMonitor trained a random forest model on FLUXNET data to estimate soil heat flux (Zheng et al., 2022). Further, ETMonitor explicitly calibrated its model using 236 flux tower sites to calibrate resistance and the VPD stress parameter for plant transpiration.

From GLEAM version 4 onwards, its transpiration stress functions for short and long vegetation were built using flux tower and sap flow sites (Koppa et al., 2022). MODIS uses flux tower data from the AmeriFlux network not only for model testing but also for model calibration where tower GPP, ET and WUE are averaged over all towers with the same biome (Mu et al., 2011).

Therefore, we think it is valid to state that locations with a higher number of flux tower sites ("data-rich regions") are likely to be more constrained when this information is also used for calibration of model parameters. The degree to which flux tower stations or their spatial density constrain ET datasets is out of scope for this work. We would also like to clarify that utilizing flux tower observations for calibration makes perfect sense and these observations are a great resource. The same information cannot really be considered an independent reference in model evaluation when it has also been used for calibration.

We will add a clarifying statement to the introduction explaining that FLUXNET observations are used for parameter calibration and machine-learning training in several of the included ET datasets, rather than solely for evaluation.

2. DCI definition (Eq. 2). Eq. 2 defines the index on significant trends only (subscript s). But Fig. 2b is described as the DCI (l. 183-184) and mapped "irrespective of significance" (Fig. 2 caption), and the text says "regardless of significance" (l. 186). Please make Eq. 2 match the way the index is used; the subscript s may be a slip.

Response: Thank you for pointing out that this can be misleading. Equation 2 is correct, as the DCI assesses the direction of significant trends relative to a selected p-value threshold. While a p-value of 0.05 is most commonly used, how one defines significance is an analysis choice. For the map in Fig. 2b, our choice was to set the p-value threshold to 1, thereby including all trends to visualise overall trend direction agreement. The approach of setting the threshold to 1 was also used in Thakur et al. (2025) to visualise overall trend direction agreement. In the topology analysis, the DCI was calculated using a p-value threshold of 0.05 to obtain a reference for the majority trend direction. However, we absolutely agree that our phrasing is misleading, and we will clarify these different applications in the revision.

3. "Onset of directional opposition" (Fig. 2d) is not defined. The metric appears only in the figure caption and in the Results (l. 208-215), and Sect. 2 does not say how it is built. The panel is hard to read as it stands: the reader has to work out the procedure from the caption, and the meaning of the dominant "≤ 1" class is not clear. Please define the metric in Sect. 2 and give the thresholds used.

Response: Thank you for pointing this out. We will definitely include this in the methodology to improve clarity. The map shows the p-value at which ET datasets begin to show opposing significant trends. In some regions, we see two datasets with highly significant trends in positive and negative directions, with p-values less than or equal to 0.01. We subsequently increase the p-value threshold for significance to 0.05, then 0.1 and finally 1. The dominant "≤ 1" class indicates that opposing trends are present but only if all trends are considered because the p-value threshold is set to 1. This class is mainly intended to distinguish these areas from areas with no directional opposition.

Editorial

- l. 74, "In this review": this is an original intercomparison, not a review.
- Figure 2 panel references in Sect. 3.1 look shifted by one: l. 192 cites Fig. 2b for the quartile disagreement, which is panel c; l. 209 cites Fig. 2c for opposition at low p-values, which is panel d. l. 186 also omits the panel letter and leaves the parenthesis open.
- Fig. 6 caption ("Where does GLEAM 4.1a oppose majority trend direction the most?") is a question fragment; please make it a descriptive caption.

Thank you for these comments that will be implemented to improve clarity. In the revised version we will follow the reviewer's suggestion for all points raised.

Reviewer 2

The manuscript entitled "Agreement, opposition, and dataset influence in global evapotranspiration trends" presents a novel framework to characterise the influence of individual evapotranspiration datasets on ensemble trend estimates. The topic is timely and relevant, and the manuscript is well written. I particularly appreciate the effort to move beyond traditional ensemble statistics and identify recurring behavioural patterns among ET datasets. I have several comments that I believe would help in further strengthen the manuscript.

Response: Thank you very much for taking the time to review our work with such rigor that will help strengthen our work.

Table 1 is highly informative. I suggest adding a column at the beginning indicating the methodological category of each dataset, together with an additional column summarising the main methodological assumptions that may influence long-term trends. This would provide valuable context for readers and would also allow the bullet-point descriptions in Lines 87–94 to be removed.

Response: Thank you for this suggestion. We will add a macro-category to the table and add a short description of each dataset's approach and remove the bullet-point description as per your suggestion. We want to avoid being speculative about the methodological assumptions behind different datasets regarding long-term trends. The lack of public information on some of the more central ET parameterization also makes a summary difficult. We would also use this as an opportunity to highlight our very detailed supplement that details the methodological approaches and identifies what information is not easily accessible.

Understanding why datasets produce the trends they do and which trends are plausible is, of course, incredibly valuable, but not supported by our study design. A further difficulty here is that we do not analyse drivers of change that could pinpoint trends to specific parameterisation, and neither do we analyse agreement of input forcing choice. One would also expect a regional dependence of what model choice/input forcing drives trends. Neither can our empirical study test the impact of specific modules. This would require a very different study design, possibly more similar to Miralles et al. (2016) who used the same

input forcing to drive different ET algorithms, or Thakur et al. (2025) who used different potential ET methods to drive the same hydrological model.

Regarding dataset selection, I encourage the authors to consider including additional widely used ET datasets. For example, PMLv2.2 has become a commonly used global dataset. Similarly, GLEAM4.1a has been superseded by GLEAM4.3a, while the GLEAM "b" version represents the most observation-constrained dataset. It would also be valuable to include thermal infrared/LST-constrained surface energy balance datasets, as well as machine-learning datasets such as FLUXCOM or X-BASE, to better represent the diversity of current global ET datasets.

Response: Thank you for your suggestions. Your paragraph contains many different points and we will address them here point by point.

We agree that GLEAM “b” is the more observation-constrained dataset, but unfortunately “b” does not cover our time period 2000-2019. This is also true for FLUXCOM and FLUXCOM-X-base. We would like to highlight that many of the present datasets employ machine learning techniques in their dataset construction. Please see the in-depth supplement for details. We do believe our dataset selection represents a wide range of datasets, and we also believe that 14 datasets are a sufficiently large number to adequately support the structured behaviour we see. Other datasets or a different period will unlikely modify the main conclusions here.

While we understand the notion to include even more datasets, the datasets we used were the ones that covered a common 20-year analysis period (with 2000-2019 as the common overlap period) at the time of conception. For every study it is necessary to include a freeze-date. Otherwise, the ensemble becomes a moving target and we would never complete our work. GLEAM was the only dataset that was updated from version 3 to 4 during the study.

Regarding your specific suggestions to include GLEAM 4.3a and PML V2.2, we would like to highlight that GLEAM 4.3a was announced after the manuscript was already submitted, and the paper led by Baez Villanueva describing the changes, such as explicitly including irrigation practices, is still under review in HESS (29.7.2026). Similarly, the final paper for PML V2.2 on ESSD is also not available (checked 29.7.2026). While exploring which datasets are available for the time period of our analysis, earlier versions of PML’s temporal coverage did not extend long enough.

On your note that it would also be valuable to include thermal infrared/LST-constrained surface energy balance, we would like to point out that ETMonitor (Zheng et al., 2022) does use such information in its algorithm for its energy balance components, but it is not an LST-residual surface-energy-balance dataset like SEBS, SSEBop, or ALEXI. These, unfortunately, also did not cover our time period. However, SSEBop was used to construct SynthesisET from 2003 onwards (Elnashar et al., 2021).

Section 2.5 requires further clarification. It is not entirely clear whether the proposed topology metrics are computed globally as single summary values or whether spatial information is retained during the calculations. A conceptual figure or a summary table describing each metric, their mathematical formulation, and its intended interpretation would

substantially improve readability. In particular, the ranking procedure and the computation of each signature should be described more explicitly. I also found the definitions of the "signal dampener" and "opposition contributor" classes somewhat difficult to interpret from Figure 1 alone.

Response: This is incredibly valuable feedback. Thank you. We will clarify our approach. We acknowledge that a flow chart would clarify the analysis steps. We wish to clarify that (1) we built a grid-cell database with each characteristic using a mathematical test (e.g. “Is the trend positive and significant?”) and therefore retained spatial information and (2) that each characteristic is defined to be binary – TRUE or FALSE. While we do describe each role, we will revisit the description and include pseudocode or test logic underneath each role to improve reproducibility and clarity. We believe that this will immediately improve understanding of “signal dampener” and “opposition contributor”.

For the signal dampener, we do the following: For each dataset and each grid cell we “test” if the trend is non-significant. We use non-significance because it is easy to measure and indicates that the variability in the time series is large compared to the trend signal and thereby “dampens” it. To compare datasets, we first sum up the areas of grid cells for which the binary test characteristic returned TRUE. We then calculate area fractions. In a third step, this area fraction is ranked to determine relative role strength. For this analysis, we only consider grid cells where all 14 datasets have valid data.

The mathematical test for “opposition contributor” is similar to a leave-one-out sensitivity. It tests whether removing a dataset would produce agreement in direction that is significant among datasets when otherwise there would be opposition in significant trends. To clarify this further, consider the example of two contrasting hypothetical grid cells. In the first grid cell, within our ensemble, datasets N1, P1 and P2 exhibit significant trends where N1 has a significant negative trend and P1 and P2 have significant positive trends. For the sake of simplicity, all other datasets have non-significant trends. Removing N1 would lead to agreement in trend direction for significant trends for that grid cell. Dataset N1 would therefore get a “TRUE” for the mathematical characteristic of “opposition contributor”. However, removing P1 or P2 would not lead to agreement in trend direction in significant trends, and they would therefore get a “FALSE” for the mathematical characteristic of “opposition contributor”. Now let us consider a second grid, where N1, N2, P1 and P2 have significant trends where N1 and N2 have significant negative trends and P1 and P2 have significant positive trends. Removing any one of these N1, N2, P1 and P2 would not lead to agreement in direction in significant trends, and they would therefore get a “FALSE” for the mathematical characteristic of “opposition contributor”.

I recommend using the most recent version of the Köppen–Geiger climate classification (Beck et al., 2023), which provides an updated global dataset.

Response: Thank you for pointing out that updated Köppen–Geiger maps exist. Our checks revealed a 6 % change in area fraction of the main Köppen-Geiger classes between the two present-day maps globally with more substantial changes in Europe and Asia. Therefore, we decided to update the panel in the Supplementary Figure using Beck et al. (2023). While the patterns remain the same and the difference is not visually distinguishable, the changes in the area fraction of quartile uncertainty categories ranged from -1% to 1.5 % across the main Köppen-Geiger classes.

The computation of the DCI considers only the direction of statistically significant trends and ignores trend magnitude. Please discuss the implications of this methodological choice, particularly in cases where datasets agree on trend direction but differ substantially in trend magnitude.

Response: The DCI was originally designed to measure the agreement in trend directions of significant trends. However, it is a methodological choice to set what one considers significant, and DCI has also been used to visualise overall trend direction agreement as we have done in Fig. 2b and as was done in other studies (Thakur et al., 2025). The DCI cannot represent the totality of trend agreement but provides a useful tool to summarise and visualise trend direction agreement. If one used the median and mean, information on how many datasets agree with that direction would be missing, and the DCI is therefore a great complementary metric to synthesise ensemble information. Of course, disagreement in magnitude can be substantial, and this is why we also included a magnitude assessment by comparing grid-cell quartile estimates and showing where quartile trend magnitudes differ substantially while also accounting for agreement in direction (Fig. 2c).

While directional disagreement is vast, we are by no means suggesting that magnitude is not important. This is why dataset-specific trend estimates across IPCC reference regions are presented in Fig. 3a, and dataset-specific maps of trend estimates are presented in the supplementary material.

Lines 184–185 state that "An index value of 1 or −1 indicates complete agreement on positive or negative trends, respectively." I found this wording somewhat misleading because the DCI as formulated in Eq. 2, only considers significant trends. Therefore, all datasets could exhibit positive or negative trends, but if only a subset is statistically significant, the DCI would not equal one. I suggest revising this statement accordingly. Additionally, DCI is defined using statistically significant positive and negative trends. However, Figure 2b appears to show DCI without considering statistical significance. This is confusing and should either be clarified or revised for consistency.

Response: Thank you for pointing this out. We will clarify this. Please see our response to reviewer 1's second comment.

The authors state that Figure 3a shows complete agreement in trend direction for only four of the 44 IPCC reference regions. Although this is an interesting result, it is difficult to identify directly from the chosen colour palette. Consider using a colour scheme that better highlights complete agreement.

Response: Thank you. We will highlight the four regions more clearly.

I suggest adding a figure summarising the upper and lower quartile uncertainty for each IPCC reference region. This would provide readers with a clearer overview of regional deviations across the ensemble.

Response: That is no problem to include. We assume that you mean trend estimates of Q25 and Q75 of the ensemble. We will add quartile trend estimates below Fig. 3a to provide a clearer visual overview of regional deviations across the ensemble.

How sensitive are the topology classes to ensemble composition? For example, would the rankings remain stable if additional datasets were included or certain datasets were removed? Similarly, are the topology rankings robust across different analysis periods?

Response: Some of the topology classes are very robust, e.g. positive signal booster, negative signal booster and signal dampener. This is because they are based on signals intrinsic to a specific dataset. Adding or removing datasets would not change the relative ranks of the remaining datasets because the mathematical binary test is dataset-specific. The three oppositional metrics are more dependent on removal or addition, because ensemble information is used to calculate the mathematical binary test.

For oppositional classes to be affected, the ensemble-based reference would have to change. In the following, we will detail the cases in which the ensemble-based reference for each role would change.

1. The "trend opposer" tests opposition to the majority trend direction using the DCI (here with a p-value threshold of 0.05) as the ensemble-based reference. The DCI-based reference could change through the addition or removal of datasets if (1) the majority trend direction were not strong and the DCI were already close to 0 and (2) one or more datasets with significant trends were added or removed. For instance, let us assume that, in a cell, three datasets have a significant positive trend and two datasets have a significant negative trend. The DCI for this specific grid cell would be $(3 - 2)/14 = 0.07$. Removing GLDAS-VIC (which often has significant positive trends) might shift the DCI from 0.07 to $(2 - 2)/13 = 0$. That cell would then have no clear majority trend direction, so all datasets would be assigned FALSE for the "trend opposer" characteristic. For the same example, if two datasets with significant positive trends were removed, the DCI would become negative, $(1 - 2)/12 = -0.083$, thereby reversing the majority trend direction. Thus, to cause a directional change in the ensemble-based reference, at least two datasets with significant trends in the same direction would need to be added or removed in cells with a weakly defined majority trend direction. Please note that we have already included a sensitivity analysis examining how the selected p-value threshold affects the topology rankings.

2. The "significance opposer" shows opposition to the majority significance level. The ensemble-based reference is a simple count of significant vs non-significant trends. In this ensemble, most cells have a non-significant majority. The metric would be affected by the addition or removal of datasets if there were no clear majority significance level because the numbers of datasets with significant and non-significant trends were equal or nearly equal. Here, the addition or removal of a dataset could then shift the majority significance level.

3. The "opposition contributor" will be most affected by adding or removing datasets because it is designed to be an outlier metric. If a dataset was added that is very similar, e.g. SEBS that is already contained in SynthesisET, SEBS might mask where SynthesisET was previously an "opposition contributor".

Here, it also becomes clear that the smaller the ensemble, the more impactful additions and removals are to the ensemble-reference metrics. In other words, this methodology requires a sufficiently large ensemble.

To showcase the stability of the topology framework, we will include a sensitivity analysis on global topology involving (1) conducting a leave-one-out experiment, (2) removing the two topmost positive signal boosters, and (3) removing the two topmost negative signal boosters.

As stated in the introduction, the period studied affects trend estimates. We will include a sensitivity analysis on how the analysis period affects the global topology rankings using two 15-year periods: (i) 2005–2019 and (ii) 2000–2014.

The topology analysis is primarily presented at the global scale. It would be valuable to extend this analysis to all IPCC reference regions to assess whether the identified behaviours are spatially consistent. Currently, only three regions (SAM, TIB, and WCE) are discussed in detail, while the remaining regional results are relegated to the Supplementary Material. I encourage the authors to consider a summary figure in the main manuscript that synthesises the topology classes across all regions.

Response: That is a nice idea. We will include a figure that shows which IPCC reference regions have a similar topology. Please note that because the IPCC reference regions and the global summary use the same grid-cell database, the global topology is an overall summary.

The Results section could be condensed without losing scientific content, as several findings are repeated across subsections.

Response: We will revisit the results section more closely to pay attention to repetitions.

The manuscript successfully classifies datasets into distinct behavioural categories but provides relatively little discussion of the physical or methodological reasons underlying these behaviours. For example, why do certain datasets consistently emerge as negative signal boosters or trend opposers? Discussing differences in forcing data, model structure, retrieval methodology, or parameterisations would considerably strengthen the manuscript.

Response: That is true, and the reasons for this are mainly discussed in our response regarding table 1. We believe this would be an interesting but different study that would test more directly what drives the trends with a regional assessment. We would also like to point out that this is not trivial. As mentioned previously, an additional complication is that documentation for many of the datasets is not always clear.

However, we will add a subsection to the discussion to highlight what conclusions can be drawn from our extensive supplementary inventory, and more importantly what cannot.

In Line 436 the authors state that the results demonstrate dataset dependence "even after harmonization to a common spatial and temporal framework." However, the temporal framework was fixed to a single analysis period rather than harmonised across multiple periods. I recommend revising this statement to avoid potential ambiguity.

Response: We will revisit this statement to avoid ambiguity and perhaps find an alternative word to summarize our pre-processing steps of regridding to a common spatial resolution and restricting the analysis to a common 20-year period.

The manuscript frequently interprets agreement within the ensemble as an indicator of robustness. However, agreement among datasets should not be confused with accuracy, particularly given the shared forcing data, model structures, and ancestry among several datasets. I encourage the authors to emphasise more clearly throughout the manuscript that

the proposed topology framework characterises agreement within the ensemble rather than dataset correctness.

Response: We agree, and we thank the reviewer for pressing on this. Agreement within an ensemble is not evidence of accuracy, and it is especially weak where members share forcing data, model structure, or ancestry. This concern has motivated a related line of work by several of the co-authors (Markonis et al., 2024), which shows that shared inputs and production algorithms can artificially narrow ensemble spread, bias the ensemble mean toward features common to related products, and obscure genuine outliers; dataset genealogy has been documented for the precipitation domain in Vargas Godoy et al. (2025). Our supplementary material including dataset information was assembled with the same concern in mind.

We will add one clarification about what the framework claims. The topology roles are defined relative to a declared ensemble: a dataset is a positive signal booster, a trend opposer, or an opposition contributor with respect to the other members present, not in any absolute sense. They are therefore descriptors of a dataset's position within an ensemble rather than measures of dataset quality, and this holds by construction rather than as a caveat.

We will address your feedback and replace "robust" and "robustness" with "ensemble agreement" or "concurrence" wherever no accuracy claim is intended. We will also clarify at the end of section 2.5 what the framework does and does not establish and add a corresponding statement in the conclusions.

Minor points:

- Extra space before the reference at Line 34.
- The statement "...ET estimation remains challenging" would benefit from additional recent references.
- Line 113: please cite the R software.
- Line 118: briefly describe how the elevation classes were adapted from Hersbach et al. (2020).
- Use en dashes consistently for numerical ranges throughout the manuscript.
- Line 184: please clarify the meaning of "pockets of negative majority trend direction."
- Use acronyms consistently after their first definition (e.g., ET, DCI, R).
- Line 307: use the spelling Köppen–Geiger consistently.
- Line 458: when discussing the contrasting conclusions of Yang et al. (2023) and Kim et al. (2021), please indicate which ET datasets and analysis periods were used, as these differences are directly relevant to the conclusions of this study.

Response: Thank you. We will include your points in the revision. Special thanks for finding an error in our data sources. Elevation classes were not adapted from Hersbach et al. (2020), but from ETOPO data. We will correct this in our revision.

References

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan,

X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.-N., 2020. The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society 146, 1999–2049. https://doi.org/10.1002/qj.3803

Koppa, A., Akash Koppa, Rains, D., Petra Hulsman, Miralles, D.G., 2022. A deep learning-based hybrid model of global terrestrial evaporation. Nature Communications 13. https://doi.org/10.1038/s41467-022-29543-7

Li, B., Ryu, Y., Jiang, C., Dechant, B., Liu, J., Yan, Y., Li, X., 2023. BESSv2.0: A satellite-based and coupled-process model for quantifying long-term global land–atmosphere fluxes. Remote Sensing of Environment 295, 113696. https://doi.org/10.1016/j.rse.2023.113696

Markonis, Y., Vargas Godoy, M.R., Pradhan, R.K., Pratap, S., Thomson, J.R., Hanel, M., Paschalis, A., Nikolopoulos, E., Papalexiou, S.M., 2024. Spatial partitioning of terrestrial precipitation reveals varying dataset agreement across different environments. Commun Earth Environ 5, 217. https://doi.org/10.1038/s43247-024-01377-9

Miralles, D.G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M.F., Hirschi, M., Martens, B., Dolman, A.J., Fisher, J.B., Mu, Q., Seneviratne, S.I., Wood, E.F., Fernández-Prieto, D., 2016. The WACMOS-ET project - Part 2: Evaluation of global terrestrial evaporation data sets. Hydrology and Earth System Sciences 20, 823–842. https://doi.org/10.5194/hess-20-823-2016

Mu, Q., Zhao, M., Running, S.W., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sensing of Environment 115, 1781–1800. https://doi.org/10.1016/j.rse.2011.02.019

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Reichstein, M., Ribeca, A., Van Ingen, C., Vuichard, N., Zhang, L., Amiro, B., Ammann, C., Arain, M.A., Ardö, J., Arkebauer, T., Arndt, S.K., Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L.B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T.A., Blanken, P.D., Bohrer, G., Boike, J., Bolstad, P.V., Bonal, D., Bonnefond, J.-M., Bowling, D.R., Bracho, R., Brodeur, J., Brümmer, C., Buchmann, N., Burban, B., Burns, S.P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T.R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B.D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P.S., D'Andrea, E., Da Rocha, H., Dai, X., Davis, K.J., Cinti, B.D., Grandcourt, A.D., Ligne, A.D., De Oliveira, R.C., Delpierre, N., Desai, A.R., Di Bella, C.M., Tommasi, P.D., Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrêne, E., Dunn, A., Dušek, J., Eamus, D., Eichelmann, U., ElKhidir, H.A.M., Eugster, W., Ewenz, C.M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., Gharun, M., Gianelle, D., Gielen, B., Gioli, B., Gitelson, A., Goded, I., Goeckede, M., Goldstein, A.H., Gough, C.M., Goulden, M.L., Graf, A., Griebel, A., Gruening, C., Grünwald, T., Hammerle, A., Han, S., Han, X., Hansen, B.U., Hanson, C., Hatakka, J., He, Y., Hehn, M., Heinesch, B., Hinko-Najera, N., Hörtnagl, L., Hutley, L., Ibrom, A., Ikawa, H., Jackowicz-Korczynski, M., Janouš, D., Jans, W., Jassal, R., Jiang, S., Kato, T., Khomik, M., Klatt, J., Knohl, A., Knox, S., Kobayashi, H., Koerber, G., Kolle, O., Kosugi, Y., Kotani, A., Kowalski, A., Kruijt, B., Kurbatova, J., Kutsch, W.L., Kwon,

H., Launiainen, S., Laurila, T., Law, B., Leuning, R., Li, Yingnian, Liddell, M., Limousin, J.-M., Lion, M., Liska, A.J., Lohila, A., López-Ballesteros, A., López-Blanco, E., Loubet, B., Loustau, D., Lucas-Moffat, A., Lüers, J., Ma, S., Macfarlane, C., Magliulo, V., Maier, R., Mammarella, I., Manca, G., Marcolla, B., Margolis, H.A., Marras, S., Massman, W., Mastepanov, M., Matamala, R., Matthes, J.H., Mazzenga, F., McCaughey, H., McHugh, I., McMillan, A.M.S., Merbold, L., Meyer, W., Meyers, T., Miller, S.D., Minerbi, S., Moderow, U., Monson, R.K., Montagnani, L., Moore, C.E., Moors, E., Moreaux, V., Moureaux, C., Munger, J.W., Nakai, T., Neirynck, J., Nesic, Z., Nicolini, G., Noormets, A., Northwood, M., Nosetto, M., Nouvellon, Y., Novick, K., Oechel, W., Olesen, J.E., Ourcival, J.-M., Papuga, S.A., Parmentier, F.-J., Paul-Limoges, E., Pavelka, M., Peichl, M., Pendall, E., Phillips, R.P., Pilegaard, K., Pirk, N., Posse, G., Powell, T., Prasse, H., Prober, S.M., Rambal, S., Rannik, Ü., Raz-Yaseef, N., Rebmann, C., Reed, D., Dios, V.R.D., Restrepo-Coupe, N., Reverter, B.R., Roland, M., Sabbatini, S., Sachs, T., Saleska, S.R., Sánchez-Cañete, E.P., Sanchez-Mejia, Z.M., Schmid, H.P., Schmidt, M., Schneider, K., Schrader, F., Schroder, I., Scott, R.L., Sedlák, P., Serrano-Ortíz, P., Shao, C., Shi, P., Shironya, I., Siebicke, L., Šigut, L., Silberstein, R., Sirca, C., Spano, D., Steinbrecher, R., Stevens, R.M., Sturtevant, C., Suyker, A., Tagesson, T., Takanashi, S., Tang, Y., Tapper, N., Thom, J., Tomassucci, M., Tuovinen, J.-P., Urbanski, S., Valentini, R., Van Der Molen, M., Van Gorsel, E., Van Huissteden, K., Varlagin, A., Verfaillie, J., Vesala, T., Vincke, C., Vitale, D., Vygodskaya, N., Walker, J.P., Walter-Shea, E., Wang, H., Weber, R., Westermann, S., Wille, C., Wofsy, S., Wohlfahrt, G., Wolf, S., Woodgate, W., Li, Yuelin, Zampedri, R., Zhang, J., Zhou, G., Zona, D., Agarwal, D., Biraud, S., Torn, M., Papale, D., 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Sci Data 7, 225. https://doi.org/10.1038/s41597-020-0534-3

Thakur, V., Markonis, Y., Kumar, R., Thomson, J.R., Vargas Godoy, M.R., Hanel, M., Rakovec, O., 2025. Unveiling the impact of potential evapotranspiration method selection on trends in hydrological cycle components across Europe. Hydrology and Earth System Sciences 29, 4395–4416. https://doi.org/10.5194/hess-29-4395-2025

Vargas Godoy, M.R., Markonis, Y., Thomson, J.R., Ballarin, A.S., Perri, S., Miao, C., Sun, Q., Hanel, M., Papalexiou, S.M., Kummerow, C., Oki, T., Molini, A., 2025. Which Precipitation Dataset to Choose for Hydrological Studies of the Terrestrial Water Cycle? Bulletin of the American Meteorological Society 106, E2000–E2016. https://doi.org/10.1175/BAMS-D-24-0306.1

Zheng, C., Jia, L., Hu, G., 2022. Global land surface evapotranspiration monitoring by ETMonitor model driven by multi-source satellite earth observations. Journal of Hydrology 613, 128444. https://doi.org/10.1016/j.jhydrol.2022.128444