Response to referee comments: Anonymous Referee #1

We greatly appreciate the reviewers for dedicating their time and offering valuable suggestions to enhance the manuscript. We have carefully revised the manuscript following the reviewers' comments. For clarity, our point-by-point responses are listed below in blue text.

Overall, I was very impressed with the written manuscript and the scientific rigor of the analysis. The introduction and methods clearly describe the complex ML, ensemble, and processed based ET estimates as well as training and validation analysis. The results showed high accuracy, and the figures were well developed and easily visualized. I have minor comments below:

Response: Thank you for your encouraging comments and constructive suggestions. We have carefully revised the manuscript to address these points and improve the overall quality of the paper.

1. Line 100: Precipitation was not used as an input covariate. Please provide an explanation for why this was not included.

Response: Thank you for your suggestion. We agree that precipitation is a primary source of terrestrial water and a vital component of the hydrological cycle. Actually, the selection of input variables in this study was based on established methodologies (Monteith, 1965; Mu et al., 2007, 2011; Penman, 1948; Priestley and Taylor, 1972; Chen et al., 2026; Shang et al., 2023; Zhang et al., 2025) and we've tested a model configuration that included precipitation as an input variable. However, we found that its inclusion did not lead to a significant improvement in model performance (Fig. R1), and therefore it was not retained in the final model. We infer that this is primarily due to the following reasons:

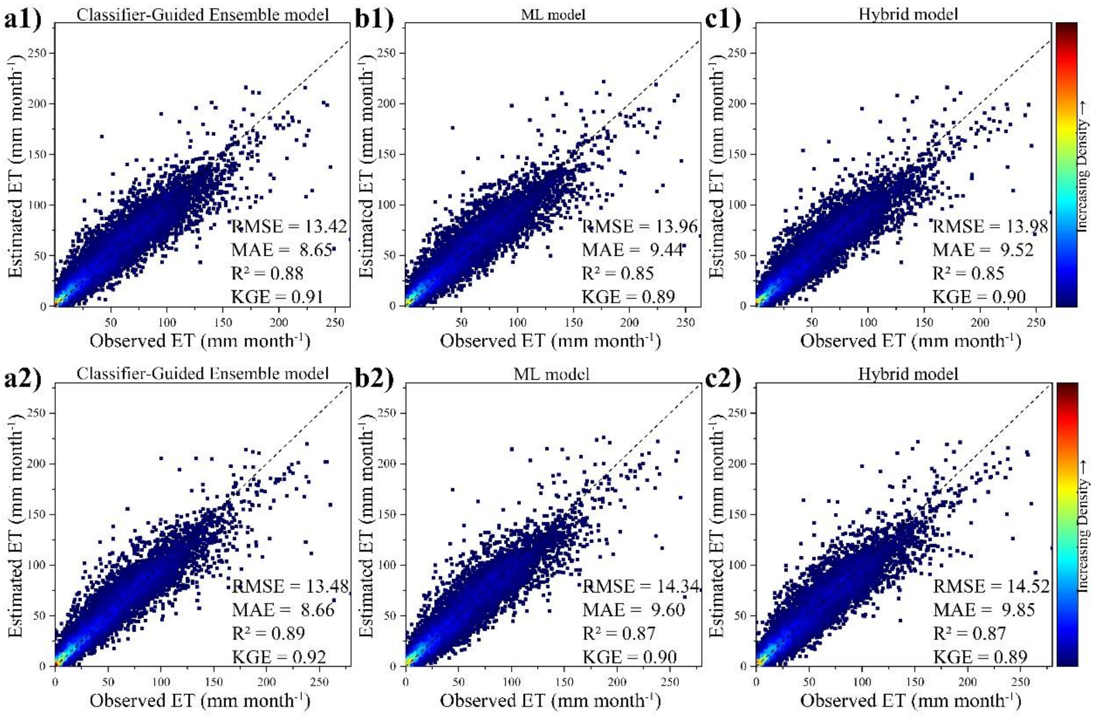

Figure R1. Performance of a) Classifier-Guided Ensemble model, b) ML model, c) Hybrid model 1) with precipitation and 2) without precipitation as an input covariate, in ten-fold cross-validation.

First, from a physical perspective, widely used ET algorithms, such as the Penman-Monteith equations (Monteith, 1965; Penman, 1948) and Priestley-Taylor equations (Priestley and Taylor, 1972), are grounded in energy balance and aerodynamic principles. In these frameworks, variables such as net radiation, vapor pressure, air temperature, and wind speed are considered the dominant driving factors, rather than precipitation (Mu et al., 2007, 2011). Notably, Vapor Pressure Deficit (VPD) incorporates the actual vapor pressure component, thereby capturing information regarding atmospheric water content. A wide range of process-based algorithms built upon these frameworks have been extensively applied (e.g., Fisher et al., 2008; Mu et al., 2007, 2011; Zhang et al., 2010), demonstrating the capability to estimate various ET components (i.e., canopy interception, vegetation transpiration, and soil evaporation) without relying on precipitation. On the other hand, soil moisture and vegetation-related variables, including the Normalized Difference Vegetation Index (NDVI), Leaf Area Index (LAI), and vegetation type, were included in the proposed model. Physically, soil moisture functions as a storage component for precipitation anomalies, while vegetation reflects the cumulative effect of water supply (Seneviratne et al., 2010; Wang et al., 2026). Therefore, these variables effectively capture the influence of changes in moisture conditions on ET without requiring precipitation as an input covariate.

Second, from a data perspective, precipitation events can adversely affect the accuracy of ET observations (Koppa et al., 2022; Medlyn et al., 2011; Shang et al., 2023). To ensure the reliability of ET estimates, we followed the methodology from previous studies (Koppa et al., 2022; Shang et al., 2023) and excluded samples collected during rainy periods. Consequently, incorporating precipitation data as an input covariate under these filtered conditions could introduce uncertainties. On the other hand, previous studies have noted that the inconsistencies among different precipitation datasets can introduce non-negligible uncertainty into ET estimation (Ma et al., 2024; Xu et al., 2020). Therefore, to avoid introducing additional uncertainty without a corresponding gain in accuracy, precipitation was excluded as an input variable.

In summary, the selected input covariate integrates not only the variables used in some widely-used process-based algorithms (Monteith, 1965; Mu et al., 2007, 2011; Penman, 1948; Priestley and Taylor, 1972), but also the high-contribution variables identified in previous studies utilizing machine learning and hybrid frameworks (Chen et al., 2026; Shang et al., 2023; Zhang et al., 2025), in which precipitation was not included as an input covariate. Consequently, the current input covariates are sufficient to achieve robust ET estimation, and the addition of precipitation would not significantly improve model performance.

Prompted by your suggestion, we realize that our original description regarding the selection of input variables was somewhat oversimplified. Therefore, we have revised the text in the Line 100 paragraph as follows:

The same input covariates for all ML model used were selected: International Geosphere-Biosphere Programme (IGBP) land cover types, leaf area index (LAI), normalized difference vegetation index (NDVI), atmospheric pressure (P), incident solar radiation (Rs), soil moisture (SM), air temperature (Ta), soil temperature (Ts), vapor pressure deficit (VPD), wind speed (WS), with a monthly temporal scale. *These variables comprehensively represent energy, water, and vegetation conditions, and have been proven effective for ET estimation in previous studies (Monteith, 1965; Mu et al., 2007, 2011; Penman, 1948; Priestley and Taylor, 1972; Chen et al., 2026; Shang et al., 2023; Zhang et al., 2025).Specifically, the selected covariates integrate not only the variables in process-based algorithms (e.g., Penman-Monteith and Priestley-Taylor equations) but also the high-contribution variables identified in established machine learning and hybrid frameworks.* The calculation process and the models used are as follows:

Reference:

[1] Chen, H., Good, S., Caylor, K., Fiorella, R. P., and Wang, L.: A hybrid Penman-Monteith and machine learning model for simulating evapotranspiration and its components, Journal of Hydrology, 134985, https://doi.org/10.1016/j.jhydrol.2026.134985, 2026.

[2] Fisher, J. B., Tu, K. P., and Baldocchi, D. D.: Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites, Remote Sensing of Environment, 112, 901–919,

https://doi.org/10.1016/j.rse.2007.06.025, 2008.
[3] Koppa, A., Rains, D., Hulsman, P., Poyatos, R., and Miralles, D. G.: A deep learning-based hybrid model of global terrestrial evaporation, Nat Commun, 13, 1912, https://doi.org/10.1038/s41467-022-29543-7, 2022.
[4] Ma, N., Zhang, Y., and Szilagyi, J.: Water-balance-based evapotranspiration for 56 large river basins: A benchmarking dataset for global terrestrial evapotranspiration modeling, Journal of Hydrology, 630, 130607, https://doi.org/10.1016/j.jhydrol.2024.130607, 2024.
[5] Medlyn, B. E., Duursma, R. A., Eamus, D., Ellsworth, D. S., Prentice, I. C., Barton, C. V. M., Crous, K. Y., De Angelis, P., Freeman, M., and Wingate, L.: Reconciling the optimal and empirical approaches to modelling stomatal conductance, Global Change Biology, 17, 2134–2144, https://doi.org/10.1111/j.1365-2486.2010.02375.x, 2011.
[6] Monteith, J. L.: Evaporation and environment., Symp Soc Exp Biol, 19, 205–234, 1965.
[7] Mu, Q., Heinsch, F. A., Zhao, M., and Running, S. W.: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, Remote Sensing of Environment, 111, 519–536, https://doi.org/10.1016/j.rse.2007.04.015, 2007.
[8] Mu, Q., Zhao, M., and Running, S. W.: 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, 2011.
[9] Penman, H. L.: Natural evaporation from open water, bare soil and grass, Proc. R. Soc. Lond. A, 193, 120–145, https://doi.org/10.1098/rspa.1948.0037, 1948.
[10] Priestley, C. H. B. and Taylor, R. J.: On the Assessment of Surface Heat Flux and Evaporation Using Large Scale Parameters, Monthly Weather Review, 100, 81–92, 1972.
[11] Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J.: Investigating soil moisture–climate interactions in a changing climate: A review, Earth-Science Reviews, 99, 125–161, https://doi.org/10.1016/j.earscirev.2010.02.004, 2010.
[12] Shang, K., Yao, Y., Di, Z., Jia, K., Zhang, X., Fisher, J. B., Chen, J., Guo, X., Yang, J., Yu, R., Xie, Z., Liu, L., Ning, J., and Zhang, L.: Coupling physical constraints with machine learning for satellite-derived evapotranspiration of the Tibetan Plateau, Remote Sensing of Environment, 289, 113519, https://doi.org/10.1016/j.rse.2023.113519, 2023.
[13] Wang, J., Xu, T., Liu, S., Kim, D., Jun, C., Bateni, S. M., Li, X., Li, X., Yang, X., Xu, Z., Zhang, G., and Ming, W.: Estimation and mechanism analysis of global evapotranspiration based on a physics-informed deep-learning model, Journal of Hydrology, 664, 134351, https://doi.org/10.1016/j.jhydrol.2025.134351, 2026.
[14] Xu, L., Chen, N., Moradkhani, H., Zhang, X., and Hu, C.: Improving Global Monthly and Daily Precipitation Estimation by Fusing Gauge Observations, Remote Sensing, and Reanalysis Data Sets, Water Resources Research, 56, e2019WR026444, https://doi.org/10.1029/2019WR026444, 2020.
[15] Zhang, C., Zhou, C., Luo, G., Ye, S., and Shi, Z.: Physics-constrained machine learning for satellite-derived evapotranspiration in China, Journal of Hydrology, 660, 133512, https://doi.org/10.1016/j.jhydrol.2025.133512, 2025.
[16] Zhang, K., Kimball, J. S., Nemani, R. R., and Running, S. W.: A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006, Water Resources Research, 46, https://doi.org/10.1029/2009WR008800,

2010.

2. Line 110: You do a great job outlining the models available with Autogluon. In line 110, you end the list with “etc.” – leading me to believe that there are even more algorithms available. In line 111, can you please list the specific ML algorithms you used?

Line 110: You say “Autogluon can combine them” – can you provide more specifics on this (perhaps just changing the phrasing) – did Autogluon combine them in your research OR autogluon can combine them – but you did not in your research. I think simply stating “Autogluon combined all the algorithms mentioned above” would suffice.

Response: Thank you for your suggestion. We acknowledge that the original description in revised Section 2.1 was ambiguous. In the revised Section 2.1, we have listed all the ML algorithms employed in this study and clarified that AutoGluon was used to combine all these algorithms to generate the final results. The revised text in the Line 110 paragraph is as follows:

Several ML algorithms are provided by Autogluon, *including CatBoost boosted trees (Dorogush et al., 2018), Extremely Randomized Trees, k-Nearest Neighbors, LightGBM boosted trees (Ke et al., 2017), Random Forests (Breiman, 2001) and neural networks.* These models have been widely used with their own distinct characteristics and advantages (Fan et al., 2019; da Silva Júnior et al., 2019; Zhangzhong et al., 2023). *Autogluon combined all the algorithms mentioned above* using methods known as stacking and bagging (Erickson et al., 2020), and can achieve better performance than individual models. More detailed algorithm information can be found in Erickson et al. (2020).

3. Figure 1: This is a great workflow figure – and helps visualize the process.

Response: Thank you very much for your positive comment.

4. Line 164: You say you used 6 well known ET products – can you please cite them after the sentence in line 164? Or perhaps say “refer to section 3.3”

Response: Thank you for your detailed suggestion. To improve clarity, we have revised the sentence in Line 164. The updated text is as follows:

At the global scale, due to the lack of reliable ET observations, we utilized six widely used global ET products, *including FLUXCOM, PLSH, GLEAM a, GLEAM b, GLDAS and ERA5-Land (refer to Section 3.3 for details),* as references to extract some relatively reliable data from the global dataset for the classification task.

5, Figure 5: Can you include the sites and land cover types of the three lowest RMSE in Line 350 paragraph. It would provide additional detail than “the majority of land covers”. I think even reference table 3 in the paragraph of Line 350 would be helpful.

Response: Thank you for your comments. Following your suggestion, we have revised the text to include specific site details. The updated text is as follows:

As Fig. 5a and Table 2 demonstrates, Classifier-Guided Ensemble model performs the best among the four models in the majority of the site, with lower average RMSE of 14.55 mm month^{-1} (Attention-Based Ensemble model: 15.41 mm month^{-1}, Hybrid model: 15.52 mm month^{-1}, ML model: 15.46 mm month^{-1}), and higher average correlation coefficient(r) of 0.90 (Attention-Based Ensemble model: 0.88, Hybrid model: 0.88, ML model: 0.87). *Specifically, the best-performing sites for multiple models include ‘CA-Obs’ (RMSE: Classifier-Guided Ensemble model: 3.92 mm month^{-1}, Attention-Based Ensemble model: 3.91, Hybrid model: 5.27 mm month^{-1}, ML model: 3.90 mm month^{-1}) and ‘CA-SF1’ (RMSE: Classifier-Guided Ensemble model: 3.75 mm month^{-1}, Attention-Based Ensemble model: 4.83 mm month^{-1}, Hybrid model: 4.08 mm month^{-1}, ML model: 5.75 mm month^{-1}). Furthermore, our model significantly outperformed the other models at ‘US-Me4’ and ‘DE-Zrk’, achieving RMSE below 4.0, whereas all other models exceeded 8.0 at these sites.*

Classifier-Guided Ensemble model also demonstrates greater performance than the other three models in the majority of land cover types (Fig. 5b and Table 3), with lower average RMSE of 16.88 mm month^{-1} (Attention-Based Ensemble model: 17.40 mm month^{-1}, Hybrid model: 17.38 mm month^{-1}, ML model: 17.94 mm month^{-1}), and higher average correlation coefficient(r) of 0.93 (Attention-Based Ensemble model: 0.92, Hybrid model: 0.93, ML model: 0.92). Classifier-Guided Ensemble model outperforms other models across most land cover types, with the exception of the CSH. *For CSH, the limitation to a single site (‘US-ZS2’) with sparse data resulted in lower accuracy for all models. In this specific case, although Classifier-Guided Ensemble model yielded better results than both the ML model and the Attention-Based Ensemble model, it did not outperform the Hybrid model.*

Further, we notice that the Attention-Based Ensemble model assigns more ‘attention’ to the ML model, since the ML model performs best at the site scale among the three base models. *Therefore, the Attention-Based Ensemble model yields results similar to the ML model across sites and land cover types, and at certain sites, such as ‘CA-SF1’, the Attention-Based Ensemble model is significantly outperformed by the Hybrid model, while Classifier-Guided Ensemble model can fully utilize the characteristics of the base models and get better results,* which is consistent with the conclusion from independent validation. In summary, the proposed model better fits the data from different sites and different land cover types, which demonstrates the effectiveness of our ensemble method.

6. Table 2: This seems repetitive and unnecessary. It is unclear how it is different than

Figure 4.

Response: Thank you for your comment. We apologize for not having clearly explained the distinct purposes of Table 2 and Figure 4. Although they may appear to both illustrate model performance across different sites, they actually assess different datasets and serve different analytical objectives. Figure 4 displays the boxplots of R^2 and KGE specifically for the 30 selected independent sites, while Table 2 summarizes the average RMSE and R^2 for sites derived from the validation set during the cross-validation process.

In addition, Table 2 is provided as a supplementary reference to Figure 5. While the Figure 5 visually compares the distribution of model performance, Table 2 provides the precise numerical metrics that are difficult to read directly from the diagram. To improve their clarity, we have revised the titles of Table 2, Table 3 and Figure 5 as follows:

Table 2. Performance of different ET models as indicated by averaged RMSE and R^2 for all sites, *based on the cross-validation dataset.*

Table 3. Performance of different ET models as indicated by RMSE and R^2 for all IGBP land cover types, *based on the cross-validation dataset.*

Figure 5. Taylor diagram that compares the performance of the four models with ground observations for a) different sites and b) different land cover types, *based on the cross-validation dataset.*

7. The overall importance of the paper is not clear. Are the global ET estimates available for public use – if so, is a link available to the dataset and what is the spatial and temporal resolution? If the estimates are available – I recommend making it clear and provide specific details and explanations about the use of the data.

On the other hand, is the paper simply a methodologic paper meant to explain a novel method for estimating global ET –although you explain the research, it would be hard for another researcher to reproduce for local or global ET estimates. If so, I recommend making it clear that the data are meant to remain proprietary, but the manuscript simply provides novel methodology.

Response: Thank you for this valuable suggestion. Yes, we would be happy to make our data publicly available. The methodology proposed in this study relies entirely on publicly available datasets and algorithms. By following the detailed steps described in the manuscript, the results can be easily reproduced to yield similar ET estimates. The dataset is generated with a spatial resolution of 0.1° × 0.1° and a temporal resolution of 1 month.

We have clarified in the Data Availability section that the global ET estimates are publicly available, and we have provided the corresponding access link: https://doi.org/10.5281/zenodo.18543737.