Dear Editor,

We thank you and the reviewers for the careful evaluation of our manuscript entitled *"Near Real-Time Estimation of Daytime and Nighttime Evapotranspiration Using GOES-R Observations and Machine Learning Models"* (MS No.: egusphere-2025-4400). We appreciate the constructive comments provided during the discussion phase.

We have carefully considered all reviewer comments and have prepared detailed, point-by-point responses addressing each concern. In our responses, we indicate how each comment has been addressed and where corresponding clarifications or revisions will be incorporated into the revised manuscript.

Overall, the reviewers raised important points regarding the demonstration of all-sky capability, uncertainty in eddy covariance observations, comparison with physically based models, temporal resolution considerations, and model generalizability. We have addressed these comments by clarifying methodological choices, expanding the discussion, and outlining additional analyses that will be included in the revised manuscript.

We appreciate the time and effort invested by the reviewers and the editorial team, and we hope that our responses adequately address all concerns raised during the review process.

Sincerely,
Sadegh Ranjbar
(on behalf of all co-authors)

RC1: 'Comment on egusphere-2025-4400', Marloes Mul, 29 Dec 2025,
https://doi.org/10.5194/egusphere-2025-4400-RC1

Opinion: Dear authors, I read your manuscript “Near real-time estimation of daytime and nighttime evapotranspiration using GOES-R observations and machine learning models” with much interest. It illustrates an interesting approach towards diurnal ET estimations for the CONUS region and it is an overall well written manuscript. I do have a few comments and suggestions for improvement as provided below.

Response: We sincerely thank the reviewer for the thoughtful and thorough review of our manuscript. We found the comments to be highly valuable and essential for improving the quality and clarity of the paper. We have carefully addressed all suggestions and believe the revisions have improved the manuscript.

Comment: Use consistent terminologies: in several figures you refer to ET estimations from EC towers as “calculated ET”, which to me is a bit confusing. In figure 2 you call it EC-derived ET, but in figure 3, the same (?) dataset is referred to as calculated ET, I think EC-derived ET is a better description.

*Response: Thank you for this helpful comment. We agree that consistent terminology is important for clarity. We have revised the manuscript to use the term **“EC-derived ET”** consistently across the entire manuscript, including all figures and figure captions, to refer to evapotranspiration estimates from eddy covariance towers that result from latent heat flux measurements. This change has been implemented to avoid confusion and to improve clarity for the reader.*

Comment: Add number of observations per station used in the annex

*Response: We have added the number of observations per station to **Table A1**.*

Comment: Line 251: you used the normalised RMSE as an indicator, an alternative is to use the relative RMSE (divided by the median or mean instead of the maximum), this indicator is less influenced by extreme values (and is also a more used performance metrics (see figure 13, Tran et al 2023).

Response: Thank you for this valuable suggestion. We agree that using a relative RMSE normalized by the median or mean is less sensitive to extreme values than normalization by the

maximum. Following this recommendation, we replaced the originally used normalized RMSE with nRMSE normalized by the median value. This choice reduces the influence of extreme values, improves robustness across conditions, and provides a more meaningful comparison, particularly for variables with strong diurnal variability and lower nighttime magnitudes. The manuscript text, figures, and tables have been updated accordingly to reflect this change.

Comment: Line 264-269 seems to fit better in a discussion section (reflection on the computation time)

Response: Thank you for this comment. We carefully considered moving Lines 264–269 to a Discussion section. However, we opted to keep this part in its current location because it strictly reports ***computational performance results*** *(i.e., training and inference time and hardware usage) without interpretation or broader discussion. Since no reflection or conceptual analysis is provided in this paragraph, we believe keeping it within the Results section helps avoid confusion and maintains a clear separation between reported results and subsequent discussion.*

Comment: Figure 2: add number of observations presented in the figure (n=..)

Response: We have revised the caption of Figure 2 to include the number of observations. Specifically, the figure now states that 17480 points are used for the daytime plot and 14304 points are used for the nighttime plot.

Comment: Table 3: what does "prediction time" mean (called prediction speed in line 238- check consistency)? Also is this result based on the calibration or validation dataset (and how is it different for validation vs calibration)?

Response: Thank you for this helpful comment. We have revised the manuscript to use the term ***"prediction time"*** *consistently throughout the text and tables. We also clarified its definition in the manuscript. The reported prediction time is for* ***the entire validation dataset****, which allows for a consistent and fair comparison between models, particularly in the context of* ***near real-time prediction applications****. We use the validation set to better reflect operational performance during real-time deployment.*

Comment: Figure 3: How is the day-time/ night-time defined? It seems the transition from day to night and night to day period is the most tricky one (and does this affect the training of the ML

and in the end the performance of the model?). Also there seems to be an overlap in the night time – started at 4PM and end 8AM (perhaps related to winter? but this is not visible in the daytime model, which should have then have included the longer evenings?). The unit is in half hour, but the graph only presents hourly data points, I would suggest to make this consistent. Caption: what does "local hour" mean?

*Response: Daytime and nighttime periods are defined using the **solar zenith angle (SZA)** rather than fixed clock hours, which allows for a physically consistent separation of day and night across seasons and latitudes. As a result, due to seasonal variations in solar geometry, some **overlap in local clock hours** (e.g., earlier night onset in winter and later night termination) is expected when data are aggregated by hour. This explains the apparent overlap in nighttime hours (e.g., 4 PM to 8 AM). The transition periods between day and night are indeed challenging; however, using SZA-based classification ensures that each data point is consistently labeled based on solar illumination conditions. This approach was applied uniformly during model training and evaluation, thereby minimizing any adverse impact on model performance. We have clarified this point in the manuscript.*

*Regarding temporal resolution, the underlying data (EC towers) are at **30-minute resolution**, while Figure 3 presents **hourly aggregated values** for clarity of visualization. We have revised the figure and caption to clearly state this to avoid confusion.*

Comment: Line 301, do you mean the one year time series is an average across all sites, or one example year for one stations or??

*Response: Thank you for this comment. The one-year time series shown corresponds to the year **2023** and represents **values averaged across all sites**, rather than a single station example. We have clarified this in the manuscript to avoid ambiguity.*

Comment: Figure 3&4, how did you calculate the daily ET, did you use the two different 'best' models? How did you deal with the transition hours (between daylight and night time)?

Response: Good question. Yes, we used two separate "best" models for daytime and nighttime ET. For Figure 3, we aggregated the half-hourly model outputs to hourly values and then averaged across all sites for each hour. For Figure 4, we combined outputs from the two models and averaged them to compute daily ET. In both cases, the underlying units remain mm hh^{-1}.

Comment: Figure 6, why did you combine certain climate classes and not include the BSh climate class? How many stations are included in each climate class?

Response: Thank you for this helpful comment. In the initial version, we focused on a subset of climate classes that were most representative based on spatial coverage in order to maintain figure clarity. Following this comment, we reconsidered this choice and expanded the analysis to include all Köppen climate classes, including BSh, to provide a more comprehensive evaluation. Figure 6 has been revised accordingly, and Figure A1 now presents the complete analysis, as including all classes in a single figure would have reduced readability and image quality. The associated text has also been updated to reflect the expanded set of climate categories. The number of stations contributing to each climate class is now reported in the revised tables and described in the text.

Comment: Figure 7, why were the other land cover classes not included?

Response: Similar to Figure 6, in the initial version, we focused on a subset of land cover classes that were most representative based on EC tower coverage in order to maintain figure clarity. Following this comment, we reconsidered this choice and expanded the analysis to include all available IGBP land cover classes to provide a more comprehensive evaluation. Figure 7 has been revised accordingly, and Figure A2 now presents the complete analysis, as including all classes in a single figure would have reduced readability and image quality. The associated text has also been updated to reflect the expanded set of land cover categories.

Comment: Line 455: the bias comment is not really substantiated. The values during the night are generally much lower than at day time and the difference may look larger in figure 3, but this is not quantified. I would suggest to include bias as a performance indicator and add it to table 3 to support this comment?

Response: In practice, the bias values in our results are generally close to zero, which is expected given the nature of the machine learning modeling. For this reason, bias did not provide strong additional discriminatory power across conditions. Following your suggestion in another comment, we instead adopted the use of ***nRMSE normalized by the median value****, which substantially improved the presentation and interpretability of the results. This metric more appropriately accounts for the lower nighttime values and quantitatively reflects the relative errors under those conditions. We have revised the analysis and tables accordingly and clarified this point in the manuscript.*

Comment: Line 515 since you are explaining the results per climate, is the comment related to the vegetation cover really relevant here?

Response: While this section primarily discusses model performance across climate classifications, we believe that referencing vegetation cover is relevant because ***climate and vegetation are inherently coupled drivers of evapotranspiration****. Differences in vegetation density, phenology, and sub-grid heterogeneity systematically co-vary with climate regimes and directly influence ET partitioning between transpiration and soil evaporation.*

Comment: Line 517-520, but would you expect that the model would perform better at these locations if it is specifically trained for those conditions?

Response: Thank you for this insightful comment. We agree that training the model specifically on sites within a given climate or environmental condition could potentially improve performance locally. However, our goal was to develop a ***generalized model applicable across diverse climates and land cover types****. The current strong performance in Mediterranean and humid subtropical climates suggests that even without climate-specific training, the model effectively captures ET dynamics where relationships among radiation, temperature, and vegetation indices are stable.*

Comment: Line 571: which result show that ALIVEet underestimates ET for 'high vegetation density and complex moisture dynamics'?

Response: Thank you for this comment. The underestimation of ALIVE ET in regions with high vegetation density and complex moisture dynamics is illustrated in ***Figure 8****, where we compare ALIVE ET with ALEXI ET. The difference map shows that ALIVE ET values are generally lower along the East Coast, and the red colors in the figure indicate areas where ALIVE ET is underestimated relative to ALEXI ET. We have clarified this point in the manuscript.*

Comment: Line 573: which result show that ALIVEet 'struggles to capture ET dynamics in the peak growing season'?

Response: The observation that ALIVE ET struggles to capture ET dynamics during the peak growing season is supported by the ***day-of-year (DOY) 149 and 179 comparisons*** *shown in Figure 8, where underestimation is most pronounced. In these comparisons, ALIVE ET shows lower values than ALEXI ET in regions with high vegetation density, indicating that the model may not fully capture the elevated ET rates typical of the peak growing season. Additionally, the* ***scatterplots in Figures 6 and 7*** *show that very high ET values deviate slightly from the 1:1 line, further indicating that ALIVE ET may not fully capture extreme ET rates.*

Edits: Comment: Line 144, remove brackets from the reference.

Response: We have removed the brackets from the reference and revised it accordingly throughout the manuscript, ensuring that the formatting of all references now aligns with the journal style.

Comment: Line 181, reduce number of names from the reference (check referencing style of journal)

Response: We have reduced the number of authors displayed in the reference to conform with the journal's referencing style and revised it consistently throughout the manuscript.

Comment: Line 194 remove initials from reference (check referencing style of journal)

Response: We have removed the initials from the reference as per the journal's style guidelines and updated the formatting throughout the manuscript.

Comment: References: Tran, B. N., van der Kwast, J., Seyoum, S., Uijlenhoet, R., Jewitt, G., and Mul, M.: Uncertainty assessment of satellite remote-sensing-based evapotranspiration estimates: a systematic review of methods and gaps, Hydrol. Earth Syst. Sci., 27, 4505–4528, https://doi.org/10.5194/hess-27-4505-2023, 2023.

Response: Thank you for suggesting this reference. We found it highly relevant and have now cited it in the manuscript to acknowledge the systematic review of uncertainty assessment in satellite-based evapotranspiration estimates.

RC2: 'Comment on egusphere-2025-4400', Anonymous Referee #2, 20 Apr 2026.
https://doi.org/10.5194/egusphere-2025-4400-RC2

Opinion: The manuscript entitled “Near Real-Time Estimation of Daytime and Nighttime Evapotranspiration Using GOES-R Observations and Machine Learning Models” by S. Ranjbar, D. Losos, S. Hoffman, Y. Zhong, J. Otkin, A. Desai, M. C. Anderson, C. Hain and P. C. Stoy presents a study with an objective to develop and evaluate a method for estimating evapotranspiration at very high temporal resolution (every 5 minutes) in near real time, using geostationary satellite observations (0.5 to 2 km) combined with machine learning models. The method intents to overcome limitations of some existing remote sensing ET products by 1) providing sub-daily estimates, for both clear and cloudy skies conditions, 2) improving nighttime reliability and, as I understand, 3) demonstrating the improvement of ET monitoring by the associated use of satellite data and machine learning in comparison with a physical model (ALEXI).
Overall the manuscript presents an interesting study with innovative elements that have a lot of potential in using such methods for near-real time monitoring. Seasonal variations are well captured for a good range of sites, daily variations seem to be well captured as well. However, the presented material and study design do not allow at this stage to convince fully about the stated objectives, as several gaps emerge and should be addressed, I think, prior to publication.

Response: We thank the reviewer for the evaluation of our manuscript. We appreciate the recognition of the novelty of the proposed framework for near real-time evapotranspiration estimation using geostationary satellite observations and machine learning. In the revised version, we have expanded the analysis, added new evaluations (including clear vs cloudy conditions, additional literature context, and extended discussion of model limitations), and clarified methodological assumptions. We believe these improvements strengthen the robustness and interpretability of the results, and we hope they address your concerns.

Comment: 1. there is a need to better demonstrate the “all-sky” capability of the algorithm, by presenting distinct statistics for clear and cloudy sky conditions, which is not very visible in the manuscript. If the authors could show that the algorithm can infer ET sub-daily variations only from remote sensing data with a good accuracy, it could be a great leap forward for the ET monitoring.

Response: We thank the reviewer for this important suggestion. The ALIVE$_{ET}$ framework is explicitly designed as an all-sky modeling system by leveraging GOES-R geostationary observations CMIs, which inherently include both clear and cloudy conditions, along with physically informed predictors such as ALIVE$_{DSR}$ and ALIVE$_{LST}$ that provide continuous surface radiative and thermal constraints independent of cloud cover.

Although our focus was daytime/nighttime evaluation of our model, clear/cloudy sky evaluation is a great idea to see the model performance under other conditions. To address this comment, we have added a new analysis (Appendix Table A3) that explicitly separates model performance under clear-sky and cloudy-sky conditions based on the GOES cloud mask classification. The results show that ALIVE$_{ET}$ maintains its predictive skill under both conditions, with only a moderate reduction in performance under cloudy skies. Importantly, the model does not rely on optical reflectance alone, but also incorporates thermal infrared channels and physically constrained variables (e.g., SZA, ALIVE$_{LST}$), enabling inference under cloudy conditions where optical-based ET models may be unable to provide predictions.

Comment: 2. there is no discussion on the possible error/uncertainty of in-situ data on the training of the algorithm or on the verification. Some major questions remain: What would be the range of uncertainty obtained if other hypotheses would have been chosen for the energy balance closure? Why a focus on nighttime ET, while in-situ data are really hard to interpret at night (noisy signal) and values very small compared to daytime? The results do not show a statistically good score ($R^2 = 0.24$), which is not supporting the stated objective.

Response: Eddy covariance measurements inherently include uncertainties arising from random turbulence sampling, and biases arising from energy balance non-closure, turbulence intermittency, and low signal-to-noise ratios, particularly during nighttime conditions when latent heat fluxes are small and often approach instrument detection limits. In our preprocessing, we applied standard quality control procedures consistent with AmeriFlux and NEON protocols, including friction velocity (u) filtering to mitigate nighttime turbulence issues. However, we did not impose additional energy balance closure corrections, as doing so would introduce assumptions that vary across sites and potentially bias cross-site consistency. Regarding nighttime ET, we acknowledge that this is a particularly challenging regime due to low flux magnitudes and higher relative uncertainty in EC measurements. In this study, nighttime ET is included to explicitly test the capability of the model to learn weak but non-zero evaporative signals driven primarily by soil heat storage release and residual canopy/soil moisture evaporation processes related to wind speed, the energy balance, xylem refilling processes and other evapotranspiration drivers.*

The reported lower R^2 values (e.g., 0.24 for nighttime with time-series LSTM) reflect this intrinsic difficulty. This study represents, to our knowledge, one of the first attempts to explicitly model near real-time nighttime ET at sub-daily resolution using geostationary satellite data and machine learning.

We have clarified these points in the revised Discussion section to better contextualize nighttime performance limitations and observational uncertainty.

Comment: 3. there is a lack of comparison of performance between the physical model (ALEXI) and the presented algorithm. The manuscript includes inter-comparison (daily and clear sky?), but not with a comparison to a reference, so it is difficult to say if it works better/worse or if it is complementary. In addition, the choice of ALEXI in that view could be motivated: why not an algorithm already estimating ET at sub-daily and continental scale, also covering cloudy sky conditions?

Response: We thank the reviewer for this important comment. We included ALEXI as a physically based benchmark because it is one of the most widely validated, operational, satellite-driven evapotranspiration models at continental scale. Unlike empirical ML models, ALEXI is based on a two-source energy balance formulation coupled with atmospheric boundary layer dynamics, enabling physically constrained ET estimation under both clear and cloudy conditions.

To strengthen the evaluation, we report spatial agreement metrics (R^2, RMSE, bias) between $ALIVE_{ET}$ and ALEXI (Figure 8) and highlight that $ALIVE_{ET}$ reproduces both spatial patterns and seasonal variability observed in ALEXI, while providing higher temporal resolution (5-min to hourly scale). This demonstrates that $ALIVE_{ET}$ is complementary to physically based models, extending their temporal resolution while maintaining consistency in spatial structure.

We have clarified this point in the manuscript to emphasize that the purpose of the $ALIVE_{ET}$–ALEXI comparison is not to determine which model performs better, but to highlight their complementary strengths and opportunities for improvement.

Comment: 4. the literature review omits other modelling systems assimilating remotes sensing data for continental scale monitoring in near-real time with physical models (e.g. Rodell et al, 2004 (global estimation); Ghilain et al, 2011 (Meteosat); and successive works), that also provide sub-daily ET estimates. Performance could be compared, either by direct inter-comparison or in a discussion.

Response: We have expanded the Introduction to include key continental-scale and near real-time evapotranspiration modeling systems that assimilate remote sensing data into physically based frameworks, including Rodell et al. (2004) and Ghilain et al. (2011), as well as subsequent developments in satellite-driven land surface modeling.

Comment: 5. the algorithm could potentially work at very high frequency (5 min), as stated in the manuscript. However, I do not find a section showing the performance of the retrieval at such temporal sampling, its reliability or its variations. Definitely, it should be worth giving statistics

at 5 min, showcasing with time series and the improvements of the statistics when aggregating to hourly and daily time scale to support the stated objective.

Response: We thank the reviewer for this critical question regarding temporal resolution. The ALIVE$_{ET}$ framework is designed to operate at 5-minute resolution, which corresponds to the native temporal sampling of GOES-R ABI observations in CONUS mode (see Losos et. al., 2025).

Losos, D., Ranjbar, S., Hoffman, S., Abernathey, R., Desai, A. R., Otkin, J., ... & Stoy, P. C. (2025). Rapid changes in terrestrial carbon dioxide uptake captured in near-real time from a geostationary satellite: The ALIVE framework. Remote Sensing of Environment, 324, 114759., https://doi.org/10.1016/j.rse.2025.114759

However, eddy covariance observations used for training are available at half-hourly resolution. To reconcile this difference, we synchronize satellite observations to the nearest EC timestamps and train the model at the native EC temporal resolution, 30 minutes. The trained model is then applied at the full 5-minute satellite sampling frequency during inference, enabling high-frequency ET estimation.

We acknowledge this introduces a scale mismatch between training and inference, and we have clarified this temporal upscaling strategy in the Methods section and added discussion on its implications for high-frequency prediction reliability.

Comment: 6. Table 3 includes a column on the time of execution of the algorithm, which is pretty much appreciated for operations. However, no comparison is done with present day physical systems (e.g. ALEXI), and no inference on the possibility to use one system or the other for near-real time, which is clearly a feature in the title, which makes difficult for the reader if it suitable for operations. It is unclear if the sentence "it remains computationally expensive" in section 3.1 means it does not fit existing near-real time operation chains.

Response: Table 3 reports inference time for ALIVE$_{ET}$ models on a standardized hardware configuration (A100 GPU for LSTM and CPU for GBR) to ensure comparability across methods. The main purpose of this analysis is to see why we chose GBR rather than LSTM. ALIVE$_{ET}$ is designed for near real-time inference at 5-minute resolution, with prediction times on the order of seconds per domain, making it suitable for operational integration into geostationary satellite processing streams. We have clarified in the revised manuscript that "computationally expensive" refers primarily to LSTM training cost rather than inference cost, and we now explicitly emphasize the operational feasibility of the GBR-based configuration for real-time applications.

Comment: 7. The study and titles aim at a continental and regional ET assessment. However, the study is limited to CONUS area using AmeriFlux and NEON sites to train the algorithm. In addition, the study reports a lower performance for evergreen broadleaf forest and savannas in several places of the manuscript. Looking at the table with AmeriFlux sites in the annex, there is neither EBF nor SAV code. Arid climates seem difficult conditions for as reliable predictions. All in all, it drives questions on: 1) the potential generalization of the algorithm as stated in the title and 2) the size and length of the sub-sampling of ground datasets. Those two aspects should be more developed in the manuscript to fill the goals of the stated objectives: Are all land cover types represented in the training set? Is it necessary? How many? Should they be in CONUS area for the algorithm to work? Could the reduced performance be related to the scarcity of the available data for training or a more difficult dynamics to grasp, as suggested in section 4.4?

Response: The training dataset includes 94 AmeriFlux and NEON sites spanning a wide range of Köppen climate classes and IGBP land cover types across CONUS. As shown in Figure 7 and Figure A2, major vegetation types such as croplands, grasslands, wetlands, and deciduous forests are represented. We acknowledge that performance is lower in sparsely represented classes (e.g., SAV, EBF), which is consistent with data scarcity rather than model structural limitations. Importantly, the model demonstrates strong transferability across well-represented ecosystems and climate regimes, indicating that performance is primarily governed by training data density rather than intrinsic model bias.

We have clarified in the revised Discussion that extending training datasets to global flux networks would likely further improve generalization in underrepresented ecosystems and is a key direction for future work.

Additional comments:

Comment: - How is trained the algorithm at 5 min sampling? Are data available at that time resolution? If not, how is the method designed to cope with different time resolution?

Response: Thank you for the comment. The models are trained using eddy covariance data at 30-minute resolution, which is the native temporal resolution of the available flux tower observations. There are no in-situ measurements at 5-minute intervals.

The 5-minute resolution refers to the inference stage, where the trained models are applied to GOES-R ABI observations available every 5 minutes. Thus, the model is trained at 30-minute scale and then driven by higher-frequency satellite inputs during prediction. We clarified this distinction in the revised Methods section.

Comment: - Table 3 reports performance metrics in comparison with EC-derived ET. Please add if it is a single site, a selection of a few sites, used or not for the training. Same for Figure 3 and 4.

Response: All performance metrics in Table 3 are computed on a site-independent test set, where 20% of EC towers are fully withheld from training. Figures 3 and 4 show aggregated results across all evaluation (test + validation) sites rather than individual stations. We have clarified this in the captions and text.

Comment: - Several time series are presented in Figure 6 and 7 (text in section 3.4), but there is no mention of the sites used or if those are averages over several sites. Are the statistics in the text section 3.4 related to single sites or more? If single sites, please mention the code name. If grouped, then you could report the variability of the scores.

Response: Figures 6 and 7 present results aggregated across all sites within each climate and land cover class, so the stats are also from more than one site. We modified the captions, and thank you for noting this as it led to improved clarity.

Comment: - An analysis of the "SHAP" values is presented in sections 3.3 and 4.3. A bit of introduction to the methodology and terminology could be done in the "methods" section to help the reader understand the results, as the formulation has not been introduced.

Response: We added a paragraph in the Methods section describing SHAP. SHAP (Shapley Additive Explanations) quantifies feature contributions to model predictions in a consistent, game-theoretic way.

Comment: - From Figure 5, does the graph for day-time suggests that DSR has less importance on ET prediction than the 4 first factors? Is it for spatial variability or temporal? I find it not very intuitive, as at 5 min intervals, DSR has been demonstrated in several studies to be a key driver of ET. A critical explanation of the graph in the view of existing references would be needed.

Response: Thank you for this observation. The lower ranking of DSR reflects feature redundancy, as radiation information is also captured by ALIVE$_{LST}$ and related thermal and vegetation variables. SHAP measures marginal contribution within the feature set, not physical importance. We clarified this interpretation in the revised Discussion.

Comment: - Figure 5 (b): it is unclear if all the drivers are shifted in time together or if only one. In addition, the axis “mean importance score” is not very intuitive and would need an explanation in the text. The text states a lower importance between 12 and 18h (I see 16-18h), that could need also a tentative explanation.

Response: All predictors are jointly shifted across the full 24-hour window. The “mean importance score” represents the average absolute SHAP value per time lag. The lower importance around mid-afternoon reflects reduced sensitivity to older states during rapid diurnal transitions.

Comment: - Figures 4, 6, 7: “calculated ET” is confusing. Are they the reference values?

Response: Thank you. “Calculated ET” refers to eddy covariance-derived ET computed from latent heat flux (LE/λ). We have revised the manuscript to use the term ***“EC-derived ET”*** *consistently across the entire manuscript, including all figures and figure captions, to refer to evapotranspiration estimates from eddy covariance towers that result from latent heat flux measurements. This change has been implemented to avoid confusion and to improve clarity for the reader.*