Response letter

Manuscript title: Learning Evaporative Fraction with Memory

We greatly appreciate the thoughtful and constructive comments of the reviewers. We have carefully revised our manuscript in response to their comments, as described in the point-by-point responses below.

The comments from the reviewers are shown in black followed and our responses in blue.

Reviewer: 2

Reviewer #2 Evaluations:

This manuscript uses machine learning models to estimate daily evaporative fraction (EF) at flux tower sites using daily precipitation, net radiation, air temperature, vapor pressure deficit, and wind speed from the tower, and daily satellite-derived LAI, along with long-term soil texture characteristics, plant functional type, annual mean tower precipitation, annual mean tower air temperature, and annual mean tower net radiation. The manuscript uses Neural Networks with a Long Short-Term Memory (LSTM) layer to encode input memory up to 365 days. Their model architecture is one 128 node LSTM layer x 1 fully-connected dense 36 node layer x 1 output dense layer w/ 1 neuron. The manuscript compares this versus a NN with 128 x 128 x 36 x 1 fully-connected nodes as a baseline.

The results indicate that the LSTM models have typical R2 values of around 0.6-0.7 for daily EF using 0-365 days of memory respectively. By training multiple models at each memory length, they find that the ensemble average estimate each day from 20 models is somewhat better than individual models (R2 of 0.64-0.73 for 0-365 days respectively). The by averaging 20 instances of the 365-day LSTM they achieve RMSE values of 0.13 across their test data set (1 tower held out per PFT). The fully connected (non-LSTM) NN has lower R2 values for individual models and for an ensemble mean at all input time series lengths.

The ensemble mean model of the 20 365-day LSTM models is shown to follow temporal patterns of daily EF dynamics across many precipitation events and periods between (Fig 3) for highlighted test sites and periods.

The manuscript further attributes predictor importance using an Integrated Gradient method, and suggests that daily precipitation, daily air temperature, and daily VPD are usually dominant predictors (when using all 365 days of input data per daily prediction), with a tendency toward temperature in more humid biomes and towards precipitation and VPD in more arid biomes.

The manuscript also breaks down the importance of more proximate versus further antecedent terms to discuss which variables' "memories" are more significant, and compare this with site rooting depth, sand fraction, and aridity index.

The manuscript concludes that their LSTM model delivers robust EF-prediction performance, adequately captures EF dynamics, that Ta and P are "the most influential driver of EF prediction", and the memory effects are a function of site characteristics (sand fraction, rooting depth, AI).

I agree with the motivating notion of the study that predicting EF (and evaporation) is difficult and poorly observationally-constrained beyond eddy covariance tower sites. I also had some concerns about the approach and conclusions drawn from the study, as well as some comments about the manuscript itself, which I detail below.

Responses:

Thank you for your positive evaluation of the motivating aspects of our study and for your thoughtful and constructive feedback, which has significantly helped us improve the manuscript. We have carefully addressed your and the other reviewers' comments in the revised version and hope our revisions satisfactorily address your concerns.

Major Points

Comment 1:

The place of Machine Learning in this study. I am not averse to the use of ML to answer science questions, particularly for data-driven prediction. In this context though, I didn't fully understand why ML was necessary to answer any specific research questions. One motivation given in the introduction was that EF is tied to soil moisture, but that soil moisture estimates are "difficult or even impossible to acquire." This is certainly true from some perspectives, but this method uses net radiation from eddy covariance towers as an input variable, which is certainly more scarcely observed than soil moisture. Additionally, while not ideal, simple indices such as

SPI are fairly decent proxies of soil moisture, accounting for "memory" processes, even with some physical justification, and can be estimated globally.

Beyond this, even completely mechanistic models of evaporation can be represented if given precipitation, VPD, Ta, wind speed, and net radiation (particularly with daily LAI, soil texture, and PFT provided as well). I recognize that observations and models do not show the same realities all the time, particularly at different spatial and time-scales, but this still strikes me as an enormous amount of input data, which really should be enough to use physical models.

Responses:

Thank you for sharing your concerns.

1. We agree that the current manuscript and introduction place considerable emphasis on the machine learning predictions. However, we would like to clarify that the core of our manuscript is to understand the drivers of EF and its memory effects through an explainable machine learning (XAI) framework. We choose Long Short-Term Memory (LSTM) as the predictive model due to their demonstrated ability to capture non-linear relationships and memory effects across time steps – capabilities that are often limited in traditional physical or other statistical models.

Considering your suggestions in Comment #4, we will revise the introduction to better clarify our motivations and research objectives.

The revised introduction is provided below. Please let us know if you have any additional concerns or suggestions - we would be happy to further refine it for improved clarity and coherence.

The revised introduction:

"*Evaporative Fraction (EF), which represents the fraction of available energy used for evapotranspiration, is defined as the ratio of latent heat flux (LE) to the sum of LE and sensible heat flux (H):* $EF = \frac{LE}{LE+H}$ *at the land surface. This ratio is critically linked to soil moisture availability, as it regulates evapotranspiration and therefore latent heat flux (Gentine et al., 2007; Nutini et al., 2014) and it also has a strong connection to vegetation greenness (Williams and Torn, 2015). Previous studies have shown that EF is mainly governed by root zone water availability and vegetation structure and is intrinsically linked to precipitation-supplied water resources (Bastiaanssen et al., 1997; Gentine et al., 2007;*

Ichii et al., 2009). As a result, EF can offer valuable insights into plant water status, and water stress, and their survival strategies during water stress conditions. A slowly declining EF during periods of droughts is indicative of a deep root system, which ensures sustained access to deep water stores (Dralle et al., 2020; Stocker et al., 2023) (Figure 1).

Current studies indicate that vegetation exhibits both resilience and resistance, collectively referred to as memory effects, during different ecosystem processes, particularly after climate extremes (He et al., 2018; Hossain et al., 2022; Canarini et al., 2021). Vegetation dynamics are shaped not only by the concurrent climate conditions but also by lagged- or memory-induced responses. For instance, favorable climate in the past may trigger vegetation overgrowth beyond the ecosystem's carrying capacity, increasing vulnerability to subsequent stress events (Zhang et al., 2021). Temperature anomalies, including heat and cold stresses, can impair vegetation function, alter its water demand (via temperature-induced soil moisture memory) and deplete carbon reserves, which in turn influence transpiration and daily EF variability (Staacke et al., 2025). Memory effects thus reflect both the capacity of ecosystem to return to equilibrium after disturbances (Xiao et al., 2024) and their ability to maintain functional stability under ongoing stress (Hossain et al., 2022; Yu et al., 2021). In Figure 1, we highlight how rooting depth, a key plant trait, mediates these memory responses to climate extremes (e.g., droughts). Despite previous efforts to explore EF mechanisms, direct observational evidence that explicitly incorporates memory effects remains limited, calling for robust analytical tools to capture and interpret the memory effect from the complex interactions between drivers.

Recent advances in machine learning (ML) algorithms have shown promise in capturing complex hydrological processes, including surface fluxes, snowpack or streamflow prediction with the effect of lagged response (ElGhawi et al., 2023; Feng et al., 2020; Jiang et al., 2022; Pan et al., 2020; Reichstein et al., 2019, 2022; Zenone et al., 2022). However, few studies have explicitly embedded memory effect in EF predictions. Long Short-Term Memory (LSTM) networks – a variant of recurrent neural networks, are particularly well suited for learning from sequential data and are increasingly used to model temporal dynamics in hydrology (Fang and Shen, 2020; Jiang et al., 2020; Li et al., 2022). LSTM, *with its recurrent cells, retains previous information from input sequences akin to how meteorological data, like precipitation, and its impact is retained over long periods of time as soil moisture or snowpack (Lees et al., 2021, 2022). In the context of vegetation, previous heat stress events can cause cellular damage that impairs photosynthesis, respiration and transpiration, ultimately altering EF (Staacke et al., 2025). The hidden states in LSTM encode system memory and evolve with time, interacting with real-time climate variables to simulate time series (Xiao et al., 2024). This allows the model to account for historical vegetation-climate interactions in a data-driven way - without relying on process-based models that often misrepresent such memory effect (Kraft et al., 2022; Lees et al., 2022). Instead, it infers such information through the memory effects captured within its recurrent cells, learning functional dependencies between EF and both current and past climate drivers such as precipitation and temperature anomalies.*

Combined with feature attribution techniques (i.e., Integrated Gradients), our proposed explainable ML framework could then extract the captured patterns and interpret the memory from "black-box" ML models. Such tools have recently led to theoretical breakthroughs in climate, ocean, and weather sciences (e.g., Tom et al., 2020; Barnes et al., 2020; Labe and Barnes, 2021), including the identifying flooding drivers (Jiang et al., 2022). In our context, interpretability enables understanding of EF's memory-driven behavior, tracking plant water stress, and even potentially inferring rooting depth (Collins and Bras, 2007; Fu et al., 2022; Liu et al., 2020; Wang et al., 2006). The identification of memory effect would not only enhance our process-based understanding of EF mechanisms, but also enable EF to serve as a more objective indicator of plant stress - one that operates independently of soil moisture estimates and is directly linked to surface evapotranspiration processes, with memory effects related to plant regulation embedded within the algorithms (Kraft et al., 2019). At the ecosystem level, difficult to have the water stress index…

In this study, we employ explainable ML to explore the memory dynamics of EF using long-term observations from 63 eddy covariance (EC) sites (>10 years) across ICOS, AmeriFlux, and FLUXNET2015 Tier 1 dataset, combined with remote sensing data. Our framework is designed to:

1) *Capture complex functional relationships between climate drivers and EF across diverse plant functional types (PFT), incorporating memory effects;*
2) *Apply integrated gradients to identify key drivers of EF predictions across different PFTs and disentangle their memory effect contributions the event-level.*
3) *Investigate how memory effects vary across diverse PFTs and relate to key site characteristics such as rooting depth, soil water holding capacity (i.e., the soil matric suction) and aridity index.*

Overall, this study advances understanding of EF regulation and memory effect. It demonstrates how explainable ML can uncover plant water-use strategies under diverse environmental regimes."

2. Regarding your concern about the use of net radiation, we have tested a version of the model that uses incoming shortwave radiation instead, which is more readily available. The model still performs well and remains suitable for our explainability-focused analyses. Please refer to our response to Comment 2 for further details.

3. As for your suggestion about using SPI, we agree that SPI may serve as a reasonable proxy for soil moisture memory effects. However, our experiments indicate that it does not match the performance of our model in capturing daily EF variability, especially the sharp changes induced by rain pulses (see our response to Comment 3). Moreover, we want to emphasize again that the core aim of our study is not just prediction but using XAI to understand how climate drivers influence EF and its memory dynamics. In this context, using the original climate forcing variables as inputs (rather than derived indices like SPI) is more informative and interpretable.

4. We acknowledge the value of mechanistic models in advancing our understanding of evapotranspiration processes. However, we would like to reiterate that our framework is specifically designed to uncover and analyze the climate memory effects embedded in EF dynamics—an aspect that is often underrepresented or misrepresented in current process-based models. By leveraging a data-driven approach with explainable machine learning, our framework enables the discovery of underlying patterns and offers new insights into the

regulation of EF, particularly by incorporating memory effects that are difficult to parameterize in traditional models.

Comment 2:

If it were the case that physical models represented EF poorly and ML methods were much better, then this might be an ideal alternative. But in this case, I was surprised by the RMSE of the models, given the wealth of input data. These models are using 20 [models] x (365 [days] x 6 [variables] + 5 [site variables]) to predict each daily EF value, along with a huge number of parameters, but the typical daily error is greater than 10%!! This makes me concerned that there are some failings somewhere in the model construction / fitting / tuning, etc.

If one used an information criterion (AIC, BIC, DIC), I would assume that it would suggest a model of EF = 0.5 as much better than the LSTM, and possibly EF = mean_annual_precip / 2000mm as better still. I don't suggest that you do this, but the manuscript needs to demonstrate that the methods are justifiable somehow.

Responses:

Thank you for raising this concern regarding model performance and the justification for using LSTM-based machine learning approaches.

As outlined in our previous responses, the primary goal of this study is not just to achieve high predictive accuracy, but to investigate the memory effects underlying evaporative fraction (EF) dynamics using an explainable deep learning framework. To this end, we opted for a time-based training/validation/test split (rather than site-based), to preserve temporal dependency and to better reflect the model's ability to learn lagged climate-vegetation interactions across time.

Using this improved data split strategy, the overall RMSE improved from 0.13 to 0.09, which will be reflected in the revised version of Figure 2.

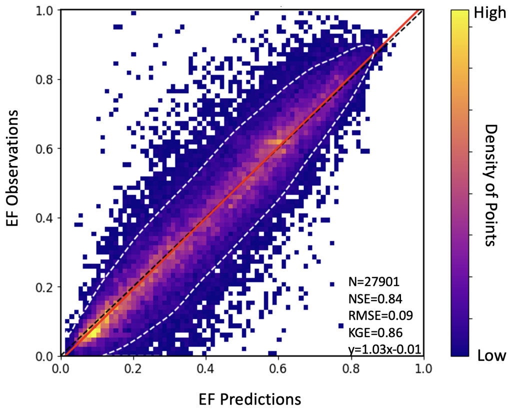

Figure R1. Model performance across all plant functional types in test set. Points in white dash line circle accounting for 90% the total samples.

We further evaluated model performance by grouping plant functional types (PFTs) into three broad categories:

- Forests: DBF, ENF, MF
- Grasslands: GRA
- Shrublands/Wooded Savanna: OSH, SAV, WSA

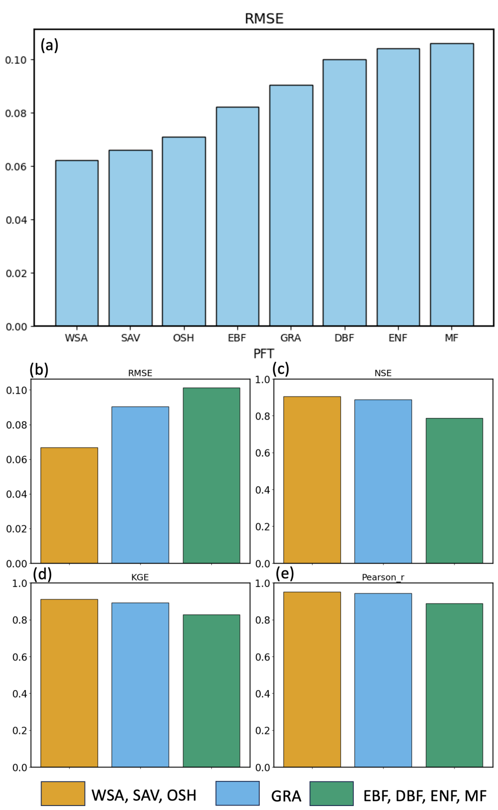

Figure R2. Model performance for each plant functional type.

We observed that forests (especially DBF, ENF and MF) tend to have slightly higher RMSE values (~0.1), likely due to their complex physiological responses and structural heterogeneity. This may be further improved by training PFT-specific models (e.g., forest-only models), and we will consider this in future work.

Importantly, we benchmarked our LSTM model against the baseline approaches you mentioned, including:

1. A EF climatology model
2. A simple SPI-based fitting model

Our LSTM-based model consistently outperformed them in capturing:

- Daily EF variability
- Dry-down dynamics
- Rain pulse responses
- Post-rainfall memory effects

(see our response in Comment 3)

These baseline models, while simple and interpretable, cannot resolve the event-level variability that is essential for analyzing memory processes in EF dynamics. Although our model's RMSE is moderate, it offers significant added value by effectively capturing temporal dependencies and the lagged responses of EF to climate drivers.

We emphasize again that the goal of our study is not merely to predict EF, but to use explainable machine learning to uncover the drivers of EF variability and their associated memory effects. This level of interpretability and mechanistic insight can only be achieved if the model is capable of learning the complex, nonlinear interactions between meteorological and ecological variables over time (please also see our revised introduction in response to Comment 1).

We will clarify these points in the revised manuscript and include additional comparisons with baseline models in the Supplementary Information to justify the modeling approach.

Comment 3:

I think the NN baseline is a fine idea to see how memory effects determine daily EF values, but only if the NN is itself convincingly good as a predictor. As a first step, I strongly suggest some simpler baseline tests:

1. I am not certain that an RMSE of 0.13 beats the daily climatology of EF for the given sites. Recognizing that we would not know the climatology of unseen sites, I would still hope that the model can do better than the seasonal cycle of EF. I would think that there could be much larger errors than 0.13 immediately following rain in arid sites, but that most days might be easily within that range.
2. A site-blind soil moisture proxy. I would also not be surprised if a model such as EF = beta0 * tanh(beta1* SPI + beta2), or something similarly parameterized, might outperform an RMSE of 0.13 as well, at least in temperate conditions. This could be fit site-by-site, biome-by-biome, or globally. This would then be a function of a single variable (P), with some memory from the SPI calculation (or any other simple weighted average).
3. Something slightly more physical, such as a simple soil moisture box model with PET-driven evaporation, particularly if using observed net radiation. This would again give the LSTM the chance to demonstrate why lagged VPD (for example) data from a year ago is a reasonable predictor.

Responses:

Thank you for your constructive and insightful suggestions. These are extremely helpful in clarifying and strengthening our manuscript.

We agree that including more baseline models is important for demonstrating the specific advantages of our LSTM-based EF model. In general, our model outperforms baseline approaches—including climatology and SPI-based models—across different plant functional types (PFTs). While baseline models may perform comparably at some temperate sites, they fail to capture key dynamics critical to our study, such as:

- Daily EF variability
- Rain pulse responses
- Post-rainfall memory effects

- Prolonged dry-down dynamics

Our main goal is not just accurate prediction, but using explainable AI (XAI) to understand the drivers of EF variability and their memory effects. These event-level insights cannot be extracted from simpler baselines, which generally miss sub-seasonal fluctuations and random rain-pulse events that drive EF at daily scales.

To address this concern, we have taken the following steps:

1. Implemented climatology models as a baseline for comparison (Figure R3).
2. Tested a SPI-based parametric model of the form $EF = \beta_0 \cdot \tanh(\beta_1 \cdot SPI + \beta_2)$.

While this model captures some degree of climate variability, its reliance on coarsely aggregated SPI values (e.g., 30-, 60-, 90-, or 365-day windows) limits its ability to detect sharp EF responses to short-term meteorological events and extended dry-down periods (Figure R3).

3. Assessed our LSTM model's performance under daily rainfall variability, demonstrating its ability to track rapid EF fluctuations that are not captured by simpler baseline models (Figure R4).

Although we believe simple soil moisture box models are useful for idealized mechanistic explorations, we think they fall outside the primary scope of our study. Our focus in the current work is to explore complex, nonlinear, and lagged interactions using explainable deep learning methods, which are better suited to uncover memory effects in data-rich, high-frequency observational records.

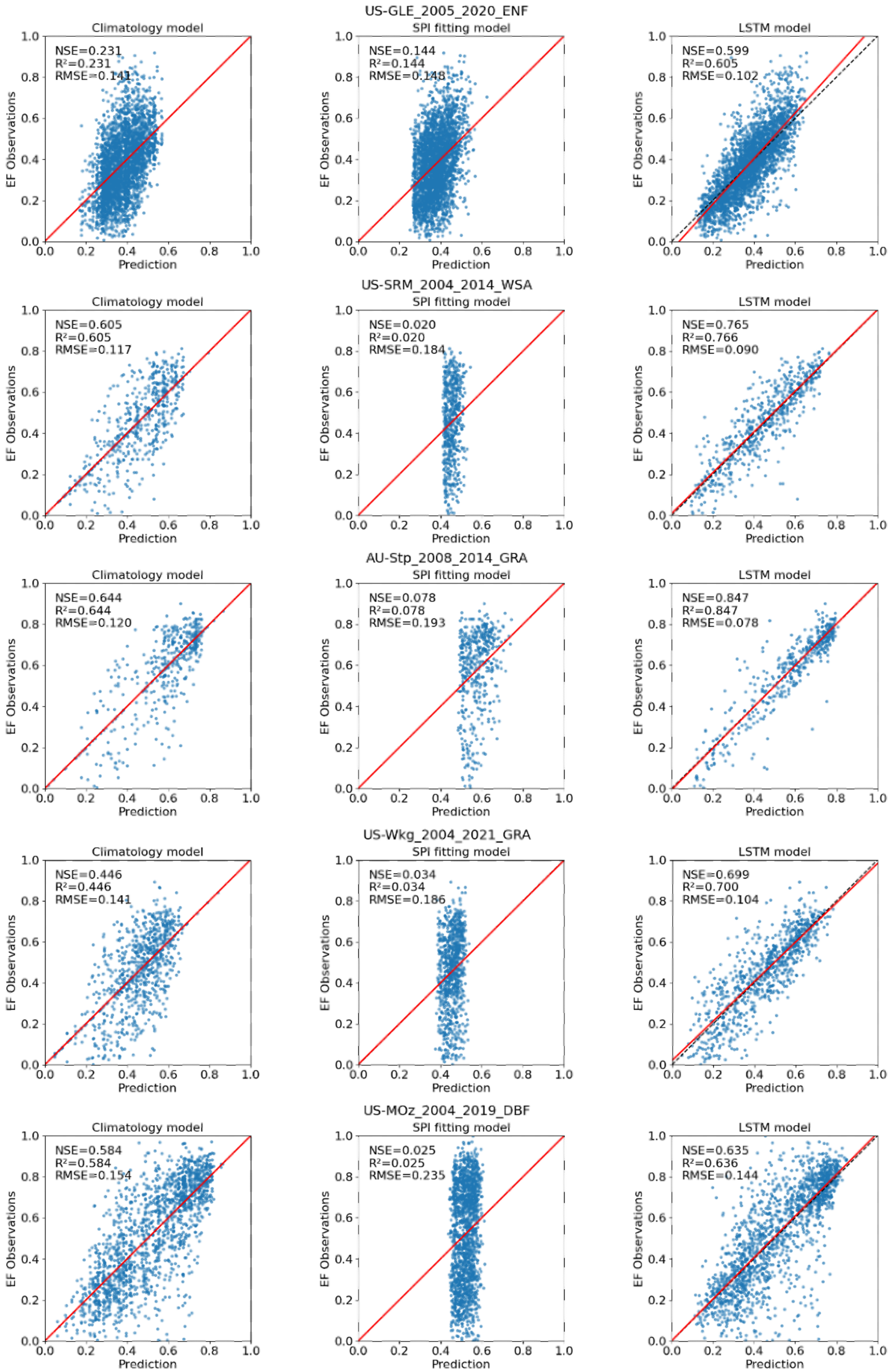

Figure R3. Comparisons between climatology, SPI-based and LSTM model.

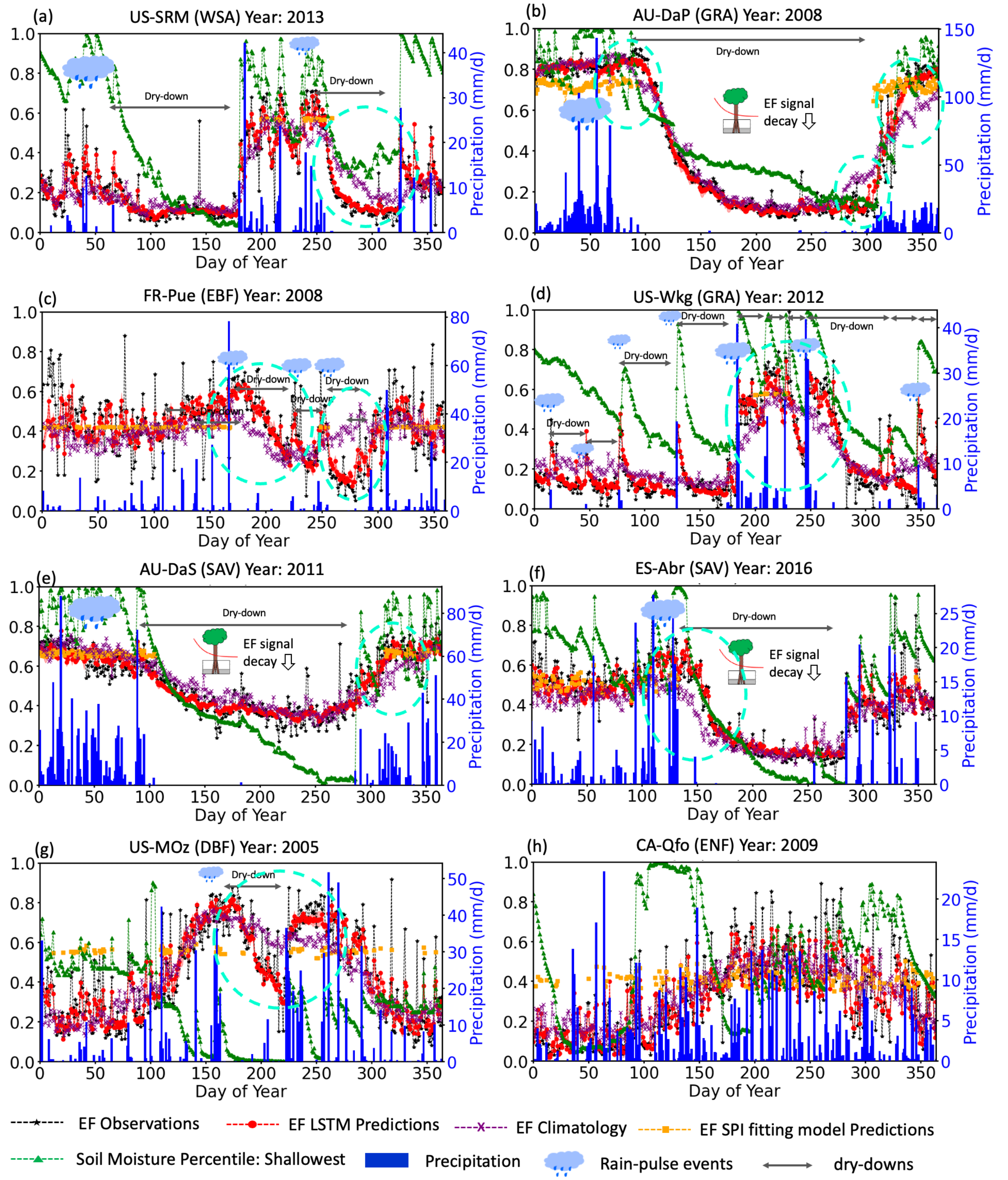

Figure R4. Comparison of evaporative fraction (EF) time series predicted by the climatology model, the SPI-based model, and the LSTM model. The LSTM model more accurately captures daily EF variability and rain pulse responses than the baseline models. The cyan dashed circles highlight instances where the LSTM model more effectively captures EF responses to rain-pulse events and the subsequent post-rain-pulse dynamics.

Comment 4:

The introduction could do a better job of motivating the idea of memory in these physical variables and why LSTM is an appropriate tool to represent it. I find it intuitive that slowly varying processes like soil moisture, which integrates precipitation and evaporation, need memory at daily scale, where they are very autocorrelated. So to some degree this is a model of soil moisture. But since the variable the manuscript is trying to predict is EF, I think a few mechanisms should be explained more deeply. For example, talk through the idea of how the last year's Ta reflects the water-demand of the atmosphere and therefore water losses from the soil (or conversely, that it represents high sensible heat fluxes and so suggests that soils have been dry for the year), and then tie that to the idea that soil moisture is the dominant predictor.

Responses:

Thank you for providing this valuable suggestion. We agree that the introduction can be strengthened by a more explicitly discussion of the role of memory in ecological systems and why LSTM is an appropriate framework to represent these biophysical processes. Your suggestions to discuss Ta memory effect can help us better clarify the motivations of our work (This is also highly relevant to Reviewer #1's comment 1).

Please check our revised introduction in Comment 1.

We will also add a dedicated paragraph in the discussions to explain why and how antecedent meteorological conditions, including soil moisture and air temperature, influence vegetation functioning and, consequently, EF. In the revised text, we review relevant literature and include evidence showing that air temperature can affect vegetation function by modulating water demand (thereby inducing temperature-related memory effects via soil moisture) and altering carbon reserves, which in turn influence vegetation transpiration and EF. The revised paragraph will read as follows:

"…Recent studies have also highlighted temperature-related memory—often referred to as heat stress (HS) memory—at the molecular scale, revealing mechanisms and regulatory layers involved in HS memory formation and resetting. These processes play a crucial role in enhancing plant stress resilience and fitness. During episodes of excessively high temperatures, plants can suffer cellular damage, primarily due to impaired photosynthesis

and respiration, accumulation of misfolded proteins, and the production of reactive oxygen species. In natural environments, plants frequently encounter multiple recurring heat stress events rather than isolated incidents. The timing of these events can vary considerably, with subsequent extremes occurring shortly after or long after the initial one (Staacke et al., 2025). This variability can significantly influence vegetation functioning and transpiration—and thereby, evaporative fraction (EF)—at later stages. Moreover, the lag time of vegetation response to high temperatures is generally longer in mid- and high-latitude regions than in low-latitude ecosystems (Xiao et al., 2024). Forests, for instance, tend to initiate resistance mechanisms earlier due to their deeper rooting systems. However, once physiological functions are affected, forests are slower to recover (Xiao et al., 2024). In contrast, grasslands, though often more vulnerable to climate extremes, typically recover more quickly (Ying et al., 2020).

Overall, the differences in rooting depth and water-use strategies across vegetation types may create a trade-off between resistance and resilience, potentially leading to post-extreme spatial heterogeneity in vegetation responses. This underscores the importance of further investigating how such memory effects—especially temperature-induced—manifest across ecosystems and influence EF…"

The citations we will add in the revised manuscript:

Staacke, T., Mueller-Roeber, B., & Balazadeh, S. (2025). Stress resilience in plants: the complex interplay between heat stress memory and resetting. *New Phytologist*, *245*(6), 2402–2421. https://doi.org/10.1111/nph.20377

Xiao, L., Wu, X., Zhao, S., & Zhou, J. (2024). Memory effects of vegetation after extreme weather events under various geological conditions in a typical karst watershed in southwestern China. *Agricultural and Forest Meteorology*, *345*, 109840. https://doi.org/10.1016/j.agrformet.2023.109840

These revisions aim to clarify the motivation for using LSTM, which is well suited for modeling temporal dependencies and lagged influences, especially in environmental systems where memory plays a key role. We hope these additions better justify our methodological choices and enhance the clarity of the study's aims.

Please also see our responses to Comments 1–3 for related discussions.

Comment 5:

The language throughout could use quite a bit of editing and proof-reading.

Responses:

Thank you. We will polish the language throughout the manuscript with the help of native speaker once the editor decided to receive our revised manuscript.

Comment 6:

The fact that the models substantially perform better when used as an ensemble mean would not be concerning for differently-parameterized or differently-structured physical models. For an ML model with this level of complexity, I have a hard time grasping why the individual models would perform so much worse. Each model has so many non-linear interacting terms and so many parameters, surely it could also account for a simple mean of the same model with tweaks to the parameters? I wonder if the training and tuning of the individual models is not as rigorous as it could be. Otherwise, you could just train another model on the input/output of the 20 models, sampled evenly, and get a single, better model. The NN training-validation-retraining process is not my area of expertise, but there is certainly literature on this.

Responses:

Thank you for raising this important point regarding ensemble modeling.

Using ensemble means of multiple machine learning models is a well-established strategy to improve generalization and reduce overfitting, particularly when dealing with complex, nonlinear systems (Dietterich, 2000). This approach has been successfully adopted in a number of recent studies (e.g., Kraft et al., 2019; Jung et al., 2019; Jiang et al., 2022; Nelson et al., 2024), including in Earth system modeling contexts similar to ours. The improved performance of the ensemble mean reflects its ability to capture complex nonlinear interactions and memory effects from high-dimensional input features—patterns that are often difficult for traditional process-based models to represent.

To further investigate this, we tested an alternative training strategy by splitting the dataset temporally (rather than by site), and observed that the difference between ensemble and individual model performance was reduced—though the ensemble mean still slightly outperformed any single model. This supports the notion that ensemble learning contributes to more robust and generalizable predictions, especially when modeling ecosystem processes influenced by antecedent climate conditions.

We appreciate the reviewer's suggestion, and will clarify this rationale in the revised manuscript.

Dietterich, T. G.: Ensemble methods in machine learning, International workshop on multiple classifier systems, 1–15, 2000.

Kraft, B., Jung, M., Körner, M., Requena Mesa, C., Cortés, J., and Reichstein, M.: Identifying Dynamic Memory Effects on Vegetation State Using Recurrent Neural Networks, Front. Big Data, 2, 31, https://doi.org/10.3389/fdata.2019.00031, 2019.

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., and Reichstein, M.: The FLUXCOM ensemble of global land-atmosphere energy fluxes, Sci Data, 6, 74, https://doi.org/10.1038/s41597-019-0076-8, 2019.

Jiang, S., Bevacqua, E., and Zscheischler, J.: River flooding mechanisms and their changes in Europe revealed by explainable machine learning, Hydrol Earth Syst Sc, 26, 6339–6359, https://doi.org/10.5194/hess-26-6339-2022, 2022.

Nelson, J. A., Walther, S., Gans, F., Kraft, B., Weber, U., Novick, K., Buchmann, N., Migliavacca, M., Wohlfahrt, G., Šigut, L., Ibrom, A., Papale, D., Göckede, M., Duveiller, G., Knohl, A., Hörtnagl, L., Scott, R. L., Zhang, W., Hamdi, Z. M., Reichstein, M., Aranda-Barranco, S., Ardö, J., Op de Beeck, M., Billesbach, D., Bowling, D., Bracho, R., Brümmer, C., Camps-Valls, G., Chen, S., Cleverly, J. R., Desai, A., Dong, G., El-Madany, T. S., Euskirchen, E. S., Feigenwinter, I., Galvagno, M., Gerosa, G. A., Gielen, B., Goded, I., Goslee, S., Gough, C. M., Heinesch, B., Ichii, K., Jackowicz-Korczynski, M. A., Klosterhalfen, A., Knox, S., Kobayashi, H., Kohonen, K.-M., Korkiakoski, M., Mammarella, I., Gharun, M., Marzuoli, R., Matamala, R., Metzger, S., Montagnani, L., Nicolini, G., O'Halloran, T., Ourcival, J.-M., Peichl, M., Pendall, E., Ruiz Reverter, B., Roland, M., Sabbatini, S., Sachs, T., Schmidt, M., Schwalm, C. R., Shekhar, A., Silberstein, R., Silveira, M. L., Spano, D., Tagesson, T., Tramontana, G., Trotta, C., Turco, F., Vesala, T., Vincke, C., Vitale, D., Vivoni, E. R., Wang, Y., Woodgate, W., Yepez, E. A., Zhang, J., Zona, D., and Jung, M.: X-BASE: the first terrestrial carbon and water flux products from an extended data-driven scaling framework, FLUXCOM-X, Biogeosciences, 21, 5079–5115, https://doi.org/10.5194/bg-21-5079-2024, 2024.

Comment 7:

I unfortunately find it difficult to evaluate the last three figures at present, until the earlier methods are either adjusted or more directly shown to be robust and high-skill.

Responses:

Thank you for sharing your concerns. We have now included all the suggested baseline models and provided a detailed comparison in the revised manuscript (see our response to your previous Comments). These additional evaluations help demonstrate the robustness and performance of our approach.

If you have any further concerns or would like to see additional experiments, we would be happy to explore those as well. We truly appreciate your constructive suggestions, which have significantly helped us strengthen the manuscript.

Minor Points

Comment 1:

I think the idea of "memory" was not very clearly laid-out. It would help to be a little more precise. Is memory the same as predictors of integrated variables? Is it the same as the idea of spectral decomposition when one is using daily data to estimate processes that aren't easy to model at daily scale? Is EF itself a process with memory, and if so, how so?

Responses:

Thank you. We have clarified the concept of memory in the revised Introduction as follows: "*Current studies indicate that vegetation exhibits both resilience and resistance, collectively referred to as memory effect, during different ecosystem processes, particularly after climate extremes (He et al., 2018; Hossain et al., 2022; Canarini et al., 2021). Vegetation dynamics are shaped not only by the concurrent climate conditions but also by lagged- or memory-induced responses. For instance, favorable climate in the past may trigger vegetation overgrowth beyond the ecosystem's carrying capacity, increasing vulnerability to subsequent stress events (Zhang et al., 2021). Temperature anomalies, including heat and cold stresses, can impair vegetation function, alter its water demand (via temperature-induced soil moisture memory) and deplete carbon reserves, which in turn influence transpiration and daily EF variability (Staacke et al., 2025). Memory effects thus reflect both the capacity of ecosystem to return to equilibrium after disturbances (Xiao et al., 2024) and their ability to maintain functional stability under ongoing stress (Hossain et al., 2022; Yu et al., 2021). In Figure 1, we highlight how rooting depth, a key plant trait, mediates these memory responses to climate extremes (e.g., droughts). Despite previous efforts to explore EF mechanisms, direct observational evidence that explicitly incorporates memory effects remains limited, calling for robust analytical tools to capture and interpret the memory effect from the complex interactions between drivers.*"

Please also refer to our response in Comment 1, where we provide additional context. We hope this addresses your concerns.

Comment 2:

Add a space between text and all citations (e.g., Line 31).

Responses:

Thank you for pointing this out. We will carefully review the manuscript and correct all instances where spaces were missing between the text and citations (e.g., Line 31), ensuring consistent formatting throughout the revised manuscript.

Comment 3:

Line 74: "…the lens of the memory effect, a perspective that becomes increasingly relevant as climate change escalates the frequency and severity of drought conditions." I'm not sure I'm following what this means.

Responses:

Sorry for the confusion. We have removed the original sentence and revised the introduction to clarify our intended message. The updated version reads:

"*Overall, this study advances understanding of EF regulation and memory effects. It demonstrates how explainable ML can uncover plant water-use strategies under diverse environmental regimes.*"