Revision Notes

Manuscript title: Learning Evaporative Fraction with Memory

We sincerely appreciate the opportunity to revise our manuscript. We are deeply grateful to the editor and reviewers for their time, thoughtful insights, and constructive suggestions, which have been invaluable in improving the quality and clarity of our work. In response to the comments received, we have undertaken comprehensive revisions to address all concerns raised. A detailed point-by-point response to each reviewer's comment is provided below, with reviewers' comments shown in black and our responses in blue.
The comments from the reviewers are shown in black followed and our responses in blue.

Reviewer: 1

Reviewer #1 Evaluations:

This manuscript applies a relatively new machine learning model to effectively capture the temporal variability of Evaporative Fraction (EF). The model shows strong agreement with observations, demonstrating its capability to represent the dynamics of EF across different PFTs and climate zones. The authors also quantitatively assess the influence of surface hydrometeorological drivers on vegetation memory, providing valuable insights into soil-plant-atmosphere interactions.

Overall, I find this study to be of interest and with potential for publication. Several aspects of the methodological description and the presentation of the results require further clarification to ensure that readers can fully understand and evaluate the work.

Responses:

Thank you for your positive evaluation and for your great efforts to help us improve our paper. We also appreciate your recognition of our study's potential to enhance the understanding of water and carbon fluxes under changing precipitation variability in the context of climate change.

In the revised version of the manuscript, we have carefully addressed your comments, as well as those from the other reviewers. Once the editor provides approval, we will upload the revised manuscript for your review.

Major Points

Comment 1:

The manuscript would benefit from a clearer and more consistent definition of “memory” throughout the text. Currently, multiple types of memory are mentioned, e.g., vegetation memory, soil memory, and meteorological memory, but the differences among these concepts are not always explicit. As a result, it can be confusing for readers to follow the main focus of the study. I recommend that the authors select one primary concept of memory and build the narrative around it, while clearly defining how other types of memory are related but distinct. This would greatly improve the conceptual clarity and help readers better understand the scope and implications of the work.

Responses:

Thank you for pointing out this.

Our primary focus is on meteorological memory as reflected in the observed daily EF variability. This type of memory, captured by the LSTM framework, implicitly integrates the influences of vegetation, soil, and atmospheric processes that contribute to the persistence of EF anomalies. We have now unified the terminology throughout the manuscript.

Comment 2:

Throughout the manuscript, the authors should explicitly highlight that this is a machine learning-based approach. At present, the manuscript briefly mentions the methodology but does not clearly differentiate it from other physical approaches for estimating EF, many of which require fewer input variables. For instance, methods such as Surface Flux Equilibrium (SFE) can also calculate EF effectively with only routine weather station data. By making this distinction clearer, the authors can better justify the novelty and added value of applying a machine learning framework, and help readers understand the specific advantages and trade-offs compared to traditional physical methods.

Responses:

Thank you.

In the abstract of our original manuscript, we have already clearly state that:

"We developed an explainable machine learning (ML) model based on a Long Short-Term Memory (LSTM) architecture, which explicitly incorporates memory effects, to investigate the mechanisms underlying EF dynamics."

Our study aims not simply to predict EF but to understand the temporal dynamics and memory effects that shape EF variability—an aspect that equilibrium-based physical models cannot capture, as they assume steady-state surface energy partitioning. Our ML-based approach maintains strong predictive accuracy even during dry-down periods, whereas the SFE framework is a non-predictive equilibrium theory that breaks down under non-stationary conditions such as dry-downs. Moreover, SFE does not explicitly incorporate precipitation inputs, while the ML model can extract and leverage precipitation-driven memory effects that are essential for predicting soil moisture decline and energy partitioning.

In addition, explainable machine learning allows us to uncover processes that are difficult or impossible to diagnose from the structural assumptions of physical models—for example, underrepresented meteorological memory effects. We hope the academic community will maintain an open-minded perspective toward machine learning and explainable AI techniques. These tools provide a principled way to peer inside otherwise opaque models and can help reveal new mechanisms—such as memory effects—that traditional physical frameworks often struggle to capture.

As suggested, we have revised the manuscript to more explicitly highlight the machine learning nature of our framework, its conceptual differences from physical EF estimation methods, and the complementary insights it provides.

Some added references:

Gharari, S., Gupta, H. V., Clark, M. P., Hrachowitz, M., Fenicia, F., Matgen, P., and Savenije, H. H. G.: Understanding the Information Content in the Hierarchy of Model Development Decisions: Learning From Data, Water Resources Research, 57, e2020WR027948, https://doi.org/10.1029/2020WR027948, 2021.

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., and Gupta, H. V.: What Role Does Hydrological Science Play in the Age of Machine Learning?, Water Resources Research, 57, e2020WR028091, https://doi.org/10.1029/2020WR028091, 2021.

Comment 3:

The manuscript would benefit from a clearer explanation of how the LSTM model improves upon previous ML approaches. At present, it is difficult for readers to fully understand why the proposed machine learning framework is needed. My own curiosity is that why earlier approaches were primarily static and, if so, why dynamic ML modeling has not been attempted before? The manuscript includes three baseline experiments, which provide a valuable opportunity to convince readers of the proposed method's strengths. I suggest the authors provide a more detailed and systematic comparison by explaining why the baselines perform worse, which would greatly enhance the clarity of the study.

Responses:

Thank you.

First, in our original introduction, we explained the specially of LSTM structure.:

"…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…"

Here we want to emphasis is earlier EF studies commonly used static or quasi-static machine-learning models (e.g., RF, MLP, GBM), which only map instantaneous meteorological forcing to EF. These models do not contain internal states or mechanisms for storing information across time, and therefore cannot capture ecohydrological memory processes such as soil-moisture carryover, precipitation dry-down progression, canopy conductance acclimation, or multi-day thermal effects that strongly modulate EF. These kind of mechanisms could be very complex and will be difficult to be represented by limited parameters and diagnostic model structures in physical models.

In contrast, LSTMs are explicitly designed to represent time-dependent processes through their gating architecture. The input, forget, and output gates allow the model to selectively retain or discard past information, enabling it to learn variable-length memory from the data. This makes the LSTM uniquely suited for capturing meteorological memory effects—especially the lagged influence of precipitation, soil moisture, and accumulated thermal conditions—compared with other ML models that lack recurrent dynamics. We now clarify these structural advantages in the revised text.

We also expanded the baseline comparison to clarify why the three baseline models perform worse. Static ML baselines fail because they cannot represent lag structures or propagate information across time. The diagnostic physical baseline, with its fixed functional forms and internally prescribed parameters, is unable to capture site-specific or event-level meteorological memory effects or their impacts on EF dynamics.

In addition, to avoid making the main text overly technical for the readers, we provide an illustration of the LSTM cell structure and update equations in the Supplementary Materials, as the reviewer suggested.

Overall, the revised manuscript now explicitly demonstrates: (i) why EF prediction requires a sequential machine learning model incorporating memory, (ii) how the LSTM's gating mechanisms provide structural advantages for learning meteorological memory, and (iii) why non-memory baselines are fundamentally unable to recover these lagged ecohydrological processes.

Kraft, B., Nelson, J. A., Walther, S., Gans, F., Weber, U., Duveiller, G., Reichstein, M., Zhang, W., Rußwurm, M., Tuia, D., Körner, M., Hamdi, Z. M., and Jung, M.: On the added value of sequential deep learning for upscaling evapotranspiration, https://doi.org/10.5194/egusphere-2024-2896, 10 October 2024.

Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, Neural Computation, 9, 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735, 1997.

Zenone, T., Vitale, L., Famulari, D., and Magliulo, V.: Application of machine learning techniques to simulate the evaporative fraction and its relationship with environmental variables in corn crops, Ecol Process, 11, 54, https://doi.org/10.1186/s13717-022-00400-1, 2022.

Han, Q., Wang, T., Kong, Z., Dai, Y., and Wang, L.: Disentangling the Impacts of Environmental Factors on Evaporative Fraction Across Climate Regimes, Journal of Geophysical Research: Atmospheres, 129, e2024JD041515, https://doi.org/10.1029/2024JD041515, 2024.

Minor Points

Comment 1:

9: What are vegetation memory effects? Please explain.

Responses:

Thank you.

As clarified in our response to Comment 1, the primary focus of this study is meteorological memory effect, as expressed through the observed daily variability of EF. We have now ensured consistent terminology throughout the manuscript.

By “vegetation memory effects,” we refer to the lagged influence of antecedent meteorological conditions on current ecosystem function. Ecosystems are shaped not only by the climate at the current time but also by past temperature, precipitation, and moisture conditions, which modulate vegetation growth, canopy conductance, soil moisture states, and energy partitioning. These delayed responses—commonly referred to as *memory effects*—can either amplify or buffer the impacts of climate extremes on ecosystem processes, including EF.

In our LSTM framework, these multi-variable meteorological memory effects are captured implicitly through the recurrent architecture, which integrates information from previous climate conditions. Although the model does not explicitly simulate physiological mechanisms, the resulting memory signatures reflect the combined influence of vegetation, soil, and atmospheric processes that govern the persistence of EF anomalies.

We now cite recent work demonstrating the strong contribution of antecedent climate to ecosystem functional anomalies:

Qiu, J., Zhang, Y., Cai, M., Keenan, T. F., Zhang, H., Gentine, P., Luo, X., Cattry, M., Zhou, S., and Piao, S.: Large contribution of antecedent climate to ecosystem productivity anomalies during extreme events, Nat. Geosci., 1–8, https://doi.org/10.1038/s41561-025-01856-4, 2025.

Comment 2:

10: I'm new to ML method, what's the difference between explainable ML and regular ML?

Responses:

Thank you.

We can add a brief explanation of the difference between explainable machine learning (XAI) and interpretable machine learning (IML) relative to regular machine learning (ML) at the first place where these terms are introduced. Specifically, we now clarify that regular ML focuses primarily on predictive accuracy, whereas XAI/IML aims to make model behavior interpretable through techniques such as Shapley additive explanations (SHAP; Lundberg & Lee, 2017), integrated and expected gradients (Sundararajan et al., 2017; Erion et al., 2021), and local interpretable model-agnostic explanations (LIME; Ribeiro et al., 2016). This addition is included to help readers—especially those without prior ML experience—better understand the methodological context of our study (Gunning & Aha, 2019; Murdoch et al., 2019; Rudin, 2022).

This clarification also highlights a key novelty of our work, as this study represents one of our first attempts to implement XAI/IML techniques to investigate evaporative fraction dynamics and meteorological memory effects.

References:

Gunning, D. and Aha, D. W.: DARPA's Explainable Artificial Intelligence (XAI) Program, AI Magazine, 40, 44–58, https://doi.org/10.1609/aimag.v40i2.2850, 2019.

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., and Yu, B.: Interpretable machine learning: definitions, methods, and applications, Proc. Natl. Acad. Sci. U.S.A., 116, 22071–22080, https://doi.org/10.1073/pnas.1900654116, 2019.

Rudin, C.: Interpretable AI: Fundamental principles and 10 grand challenges, Nat. Mach. Intell., 4, 611–619, https://doi.org/10.1038/s42256-022-00534-z, 2022.

Lundberg, S. M. and Lee, S.-I.: A unified approach to interpreting model predictions, in: Advances in Neural Information Processing Systems, 30, 4765–4774, 2017.

Sundararajan, M., Taly, A., and Yan, Q.: Axiomatic Attribution for Deep Networks, http://arxiv.org/abs/1703.01365, 2017.

Erion, G., Janizek, J. D., Sturmfels, P., Lundberg, S. M., and Lee, S.-I.: Improving performance of deep learning models with axiomatic attribution priors and expected gradients, Nat. Mach. Intell., 3, 620–631, https://doi.org/10.1038/s42256-021-00343-w, 2021.

Ribeiro, M. T., Singh, S., and Guestrin, C.: "Why Should I Trust You?" Explaining the Predictions of Any Classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144, https://doi.org/10.1145/2939672.2939778, 2016.

Comment 3:

11: Should be "vegetation memory effects"

Responses:

Thank you. We have now unified the terminology throughout the manuscript by consistently using "memory effects" after explicitly defining this term as referring to meteorological memory effects in the revised Introduction. Please also check our responses to your minor comment 1.

Comment 4:

14: What's the advantage of this study compared to SFE (Surface Flux Equilibrium), which also only require routine weather station data for EF calculation?

Responses:

Thank you.

The Surface Flux Equilibrium (SFE) theory provides a physically grounded framework to diagnose the equilibrium relationship between evaporative fraction and meteorological states (e.g., air temperature, humidity, wind speed). However, SFE inherently assumes quasi-stationary surface flux equilibrium and does not represent temporal evolution or memory effects. Therefore, while SFE-based EF estimates are appropriate for long-term or climatological analyses (e.g., monthly or seasonal averages), they cannot capture high-frequency EF dynamics and delayed responses to meteorological forcings, especially the dry-downs. In contrast, our LSTM framework explicitly incorporates temporal dependencies, allowing it to learn non-equilibrium, memory-driven adjustments in EF that arise from soil moisture persistence, canopy heat storage, and vegetation physiological lag.

Please also check our responses to your major comment 2.

Comment 5:

17-19: Which corresponds to "water-limited" and "energy-limited" regimes?

Responses:

Thank you.

In Lines 17–19, our intention was not to explicitly classify sites into "water-limited'' or "energy-limited'' regimes, but rather to summarize the dominant drivers of EF dynamics within each vegetation type. When incorporating memory effects, precipitation and VPD—reflecting atmospheric dryness—emerge as primary controls in woody savanna, savanna, open shrubland, and grassland sites. In contrast, in most forest sites (which are typically less water-limited), air temperature dominates EF variability, while net radiation primarily captures instantaneous effects.

In general, savanna and shrubland ecosystems tend to operate under water-limited conditions, whereas forests are more often energy-limited. However, as shown in Fig. 7, substantial site-level variability exists within each PFT.

We agree that explicitly grouping sites into water-limited versus energy-limited regimes could provide additional insight. If requested by the Editor, we are open to expanding the discussion to incorporate this perspective in a future revision.

Comment 6:

24: Previously it says "vegetation memory effects", please be consistent.

Responses:

Thank you. As you suggested, we have ensured consistent use of the term throughout the manuscript.

Comment 7:

31: I think here you don't have to emphasize root-zone, since SM-EF at surface soil layer should be stronger.

Responses:

Thank you.

We have revised the text to use "soil water availability" instead of emphasizing root-zone: "Previous studies have shown that EF is mainly governed by soil 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)."

Comment 8:

36: I feel the cause-and-effect of this paragraph should be rephrased as how vegetation memory influence EF, rather than the other way around. The goal of this study is to predict EF, and vegetation memory is one of the key drivers. Or you should place this content after the description of EF prediction.

Responses:

Thank you.

We want to clarify that our goal is not mainly predict EF but our first aim of this study is to implement LSTM based explainable machine learning which incorporating memory effect to understand the EF dynamics. Due to the understanding task, we need an EF prediction model that can better capture the EF dynamics not only during dry-downs but also non-dry-downs.

Here we thought it's necessary to mention that this is actually a second-round review of our manuscript. In the previous review round, we have the introduction more focus the EF predictions and placing the part of memory effect following the description of EF prediction: Attach part of the previous version of introduction of EF predictions:

"…Accurate prediction of EF is essential for estimating evapotranspiration, monitoring plant water stress, and potentially inferring rooting depth (Collins and Bras, 2007; Fu et al., 2022; Liu et al., 2020; Wang et al., 2006). Among various environmental factors affecting EF, the soil moisture state dominates the day-to-day EF variations, particularly in water-limited regions (Dirmeyer et al., 2000; Dong et al., 2022; Haghighi et al., 2018; Lu et al., 2016). It serves as both a direct predictor of EF in prolonged dry spells and as a historical record, mirroring past climatic conditions and water availability — analogous to a memory effect. However, precise measurement or estimation of soil moisture in the rooting zone or plant water stress, is difficult or even impossible to acquire (O. and Orth, 2021; Skulovich and Gentine, 2023). This led to the development of various drought indices such as the Palmer Drought Severity Index, which try to infer the changes in soil moisture through a mechanistic, yet empirical, approach. These constraints underscore the importance of a more direct predictive model for EF, which could in turn be used as a more objective stress index, that can operate independently of soil moisture estimates and directly linked to surface evapotranspiration, emphasizing the incorporation of memory effects related to plant regulation within the algorithms (Kraft et al., 2019)…."

However, reviewers pointed out that simpler models can achieve EF prediction accuracies comparable to the LSTM, and therefore encouraged us to shift the introduction away from predictive performance and toward highlighting the memory effect, its underlying mechanisms, and how an LSTM-based explainable ML framework can extract these functional relationships. We revised the introduction accordingly. This restructuring better reflects the true focus of our study, as the majority of our analyses center on Expected Gradients–based driver interpretation rather than on EF predictive performance. However, we have also improved the flow of the Introduction so that the discussion of memory effects aligns more naturally with the subsequent description of the EF prediction framework.

Comment 9:

42: Please explain the terminology the first time it appears.

Responses:

Thank you. We have revised this.

"… via temperature-induced soil moisture memory, defined as the persistence of soil moisture anomalies driven by past temperature conditions through their effects on evapotranspiration and recharge dynamics …"

Comment 10:

74: Again, what's the difference between "explainable ML" and regular ML method?

Responses:

Thank you. We have revised this.

Please check our response to your minor comment 2.

Comment 11:

115: Did you also mask those with energy imbalance larger than a threshold (e.g., (Rn-Gs)-(LH+SH)>30W/m2)? Please explicitly indicate it in the main context.

Responses:

Thank you.

In this study, we used the energy-balance-closure–corrected fluxes provided in the FLUXNET dataset, namely LE_CORR and H_CORR, which adjust the raw latent and sensible heat fluxes to satisfy Rn - G ≈ LE + H following the standard procedure in Wilson et al., 2002. Because these corrected fluxes already account for energy imbalance, we did not apply an additional masking based on an imbalance threshold such as (Rn - G) - (LE + H) > 30 W/m2. We now explicitly clarify this in the main text and the supplementals.

In addition, as noted in the Methods section, “We assume that the errors in LE and H have comparable magnitudes, consistent with previous studies (Foken, 2008; Hollinger and Richardson, 2005; Richardson et al., 2006), and are uncorrelated. This assumption allows the mathematical elimination of errors associated with the lack of energy balance closure (Schwalm et al., 2010).”

Wilson, K., Goldstein, A., Falge, E., Aubinet, M., Baldocchi, D., Berbigier, P., Bernhofer, C., Ceulemans, R., Dolman, H., Field, C., Grelle, A., Ibrom, A., Law, B. E., Kowalski, A., Meyers, T., Moncrieff, J., Monson, R., Oechel, W., Tenhunen, J., Valentini, R., and Verma, S.: Energy balance closure at FLUXNET sites, Agricultural and Forest Meteorology, 113, 223–243, https://doi.org/10.1016/S0168-1923(02)00109-0, 2002.

Comment 12:

121: What is "corrected" LH? Please explain.

Responses:

Thank you.

In this study, we used the energy-balance-closure–corrected fluxes provided in the FLUXNET dataset, namely LE_CORR and H_CORR, which adjust the raw latent and sensible heat fluxes to satisfy Rn - G ≈ LE + H following the standard procedure in Wilson et al., 2002.

Please also see our response in your minor comment 11.

Comment 13:

Figure 4. Please indicate the number of sites included for each PFT.

Responses:

Thank you.

Figure 4 is an example time-series plot showing selected sites and years for each plant functional type (PFT). We have revised the Figure 4 caption to explicitly indicate that this figure presents representative sites rather than all sites.

"Figure 4. Example time-series plots for representative sites and years for each plant functional type…"

In addition, the number of sites included for each PFT is clearly reported in the summarized results shown in Figure 3 and Figure 6, where the site counts are provided in parentheses following each PFT label (e.g., DBF (19), ENF (24), GRA (18), etc.).

Comment 14:

143: Section 3.2, I suggest the authors describe explicitly the difference of each baseline model from LSTM. To non-expert in ML, it looks like the settings of FNN is similar to LSTM, then why does FNN perform worse? Additionally, why do the authors want to add the SPI-based model, and why the SPI model performs so bad ($R2<0.1$)? Please add relevant discussions.

Responses:

Thank you.

We have added explanations in Section 3.2 to clarify the conceptual differences among the baseline models and to discuss why the feedforward neural network (FNN) and the SPI-based model perform substantially worse than the LSTM.

First, although the FNN uses the same meteorological time-series inputs as the LSTM and has a comparable number of parameters, it differs fundamentally in structure. The FNN treats all input features as independent and lacks any mechanism to retain information from previous time steps. In contrast, the LSTM is explicitly designed to capture temporal dependencies through gated recurrent units, allowing it to learn antecedent-condition effects that are essential for modeling EF during dry-down periods. This structural limitation explains why the FNN cannot capture memory effects and therefore underperforms relative to the LSTM.

Second, we included the SPI-based model because a reviewer in the previous round requested an additional simple baseline beyond the FNN. SPI is a widely used hydrological drought index and provides a transparent reference grounded in traditional drought assessment methods.

Attached the reviewers' comments from the previous review round:

"Reviewer #2 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:
>
> …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)…"

Finally, the poor performance of the SPI-based model ($R^2 < 0.1$) is expected because SPI summarizes only precipitation anomalies over aggregated windows (e.g., 30–365 days) and does not incorporate key EF drivers such as radiation, temperature, vegetation state, wind, or short-term soil moisture dynamics. As a result, SPI cannot capture rapid EF fluctuations following rain pulses, nor the nonlinear meteorological memory effects that the LSTM model is designed to resolve.

These clarifications have been added to the revised manuscript.

Comment 15:

145: In figure 3 it says FNN, please be consistent.

Responses:

Thank you. We have unified to FNN.

Comment 16:

176: I suggest the authors also add a brief description of EG in the main context, since it is an important component of this study.

Responses:

Thank you. We can add the brief description of EG in the main context.

Comment 17:

212: -1.21, Is it a typo? Why R2 can be negative and larger than 1?

Responses:

Thank you. This was a typo. We have corrected R^2 to the Nash–Sutcliffe Efficiency (NSE) coefficient. Unlike R^2, NSE can take values less than 0, which explains the negative value previously shown. The revision has been updated in the manuscript.

Comment 18:

216: “Figure 4” should be “Figure 3”?

Responses:

Thank you. We have revised this.

Comment 19:

293: From here to 299, can be moved to method part.

Responses:

Thank you. We will move it to the method part.

Comment 20:

302: Please indicate this is Shortwave radiation (RAD)

Responses:

Thank you. We reivsed it as suggested.

Comment 21:

307-309: Isn't air temperature correlated with radiation?

Responses:

Thank you for pointing this out. Yes, air temperature and radiation are indeed correlated (mean $R^2 = 0.35$ across our sites for the whole period), and we have revised the relevant sentences and added statements acknowledging this limitation. Nevertheless, it is important to note that most studies include both air temperature and radiation when predicting ET, despite their correlation. We note that the Ta–radiation correlation varies systematically across climate regimes: correlations tend to be lower in arid sites, where air temperature is strongly influenced by turbulent mixing and sensible-heat–dominated surface conditions, and higher in humid sites where temperature is more directly constrained by available radiation. This spatial variability further highlights that Ta contains distinct information from radiation, especially in water-limited ecosystems.

In addition, in our context, radiation primarily governs short-term EF variability through instantaneous energy input, whereas air temperature exerts a stronger influence over longer timescales when memory effects are considered. This distinction justifies their separate contributions in the model. Although future work using causal-inference frameworks could more rigorously disentangle their individual effects on EF dynamics, such an analysis is beyond the present scope. We have added this point to the limitations section.

Comment 22:

351: Figure8, Can you re-arrange this figure with x-axis ranging from shallow to deep rooting-depth? Plus, how do you normalize the Contributions? Why are contributions from all variables even lower than precipitation alone? Does this indicate there could be negative feedbacks between variables?

Responses:

Thank you.

We have revised Figure 8 as suggested.

Regarding the normalization of contributions, we would like to clarify that the contributions shown in Figure 8 are based on the absolute values of Expected Gradients (EGs). EGs can be positive or negative depending on whether a variable at previous timesteps increases or decreases EF; however, our goal here is to quantify the magnitude of memory effects rather than their directional influence. Therefore, for each PFT we compute:

$$PC(S,t) = \frac{\sum_{i \in S} \sum_{t=k}^{t=365} \left| EG_{i(t)} \right|}{\sum_{i \in S} \sum_{t=0}^{t=365} \left| EG_{i(t)} \right|} \times 100\%$$

In this formula, S denotes the set of variables whose memory contributions are evaluated (precipitation only in Fig. 8a, and all dynamic predictors in Fig. 8b). The index i in S sums over the variables in this set. The term $EG_i(t)$ is the Expected Gradient attribution for variable i at lag t, and its absolute value is used to quantify the magnitude of memory. The variable t represents the time lag in days, and k defines the lower bound of the antecedent window, ranging from 7 to 175 days for different boxplots in Figure 8. The numerator accumulates memory contributions from lag k to 365 days, whereas the denominator sums contributions from lag 0 to 365 days. Thus, the ratio measures the percentage of total memory contributions attributable to lags deeper than k days for a given variable set S.

To generate Figure 8a and 8b, we evaluate this metric using two different variable sets: S = {P} for precipitation-only memory, and the full predictor set S = {P, Ta, RAD, VPD, WS, LAI} for the "All Variables" case. No other aspect of the calculation differs between the two panels.

This normalization ensures that all values are positive and comparable across variables and prevents positive and negative EG attributions from canceling each other out. Accordingly, the fact that precipitation exhibits a higher long-term contribution than the aggregated "All Variables" case does not imply negative feedbacks. Instead, it reflects two factors. First, precipitation has a substantially longer effective memory than other drivers, consistent with ecohydrological theory. Second, when all variables are included, the memory contributions decreases (compared to the precipitation-only memory effect) because the normalization is performed over a larger set of predictors, many of which (e.g., shortwave radiation) exhibit much shorter memory.

To avoid ambiguity, we have added explicit explanations in the Methods section describing (i) how EG-based memory contributions are computed and (ii) why precipitation alone exhibits a larger apparent long-term contribution than the aggregated "All Variables" case.

Comment 23:

356: Should be "temporal EF machine learning model".

Responses:

Thank you. We have revised it as suggested.