Revision Notes

Manuscript title: Learning Evaporative Fraction with Memory
Manuscript number: egusphere-2025-4082

We thank the editor and reviewers for their positive assessment and constructive suggestions. We have addressed all remaining comments below. Reviewer comments are shown in black and our responses in blue.

Response to the Editor

Editor decision:

Dear Authors, Three reviewers evaluated your revised manuscript very favorably. However, two of the reviewers suggested a few minor additional changes. I would appreciate it if you could review their comments and make the corresponding revisions. If you disagree with any of the suggestions, please clearly state your rebuttal in your response.

Responses:

Thank you for the favorable decision. We have addressed all remaining comments from Reviewers 2 and 3 and revised the manuscript and Supplement accordingly.

Reviewer: 1

Report outcome:

The referee recommended acceptance as is and provided no further written suggestions for revision.

Responses:

We thank the referee for the positive evaluation and recommendation.

Reviewer: 2

General evaluation:

Overall, I appreciate the authors' efforts in revising the manuscript. They have addressed the previous review comments well, and I continue to think that this is an innovative approach that provides valuable insights into ecosystem memory effects. Nevertheless, after reading the revised version, I encountered several aspects of the presentation and interpretation that should be resolved before the manuscript is suitable for publication.

Responses:

We thank the reviewer for the positive assessment. The remaining presentation and interpretation issues have been addressed as detailed below.

Comment 1:

First, I recommend replacing the term "PFT" with "vegetation type" throughout the manuscript. The analyses are performed at the ecosystem level, where the diagnosed fluxes inherently reflect the influence of species composition rather than individual plant functional types. This terminology would therefore be more appropriate. However, this change also complicates the comparison with the rooting depth data from Tumber-Davila, which are reported at the species level. The manuscript should explicitly discuss this mismatch in scale and its implications for interpreting the observed relationships.

Responses:

Thank you. The eddy-covariance fluxes and the derived memory metrics represent ecosystem-scale responses within each tower footprint and may integrate multiple species. We nevertheless retained “PFT” where it refers to the standard IGBP classification assigned to each flux site and is used as a site-level attribute or grouping label familiar to the flux-tower community. We thus clarified the limitation of this definition (it represents only the dominant vegetation around the tower) and used vegetation-type or ecosystem-level terminology in the ecological interpretation where appropriate, particularly when comparisons are made with the species-level rooting-depth observations in Figure 8.

The mismatch mainly arises in the rooting-depth comparisons. The flux-derived metrics are ecosystem-scale, whereas the Tumber-Dávila et al. (2022) rooting-depth observations are species-level. We therefore regrouped the memory and rooting-depth comparisons at the biome level following Stocker et al. (2023) and Olson et al. (2001). No conclusions are changed. The revised manuscript and Figure 8 also explicitly acknowledges the spatial mismatch, substantial within-PFT variability, and the coarse representation of vegetation heterogeneity by IGBP classes. Accordingly, these relationships are interpreted as broad correlative indicators rather than direct quantitative estimates of rooting depth (Sections 2.1, 3.6, 4.3, and 5.2; Figures 8 and 9a).

Comment 2:

Second, I found the relationship between the different training and evaluation strategies and the underlying research question to be unclear. The results from the different approaches are presented alongside one another, giving the impression that they are interchangeable, whereas they address fundamentally different questions. Since the primary objective of the study is diagnostic rather than predictive, it would seem more appropriate to focus on site-specific models trained and evaluated using temporal cross-validation (which I believe corresponds to the authors' option 2). At a minimum, the rationale for including multiple strategies and the interpretation of their respective results should be explained much more clearly.

Responses:

Thank you. The four training and evaluation strategies were included both as benchmarking analyses for distinct generalization questions and to address the concern raised in the previous review round regarding the high reported model performance. They are not treated as interchangeable approaches. The time-based split and time-split cross-validation evaluate temporal generalization within observed sites, whereas site-split cross-validation and LOSO evaluate spatial transfer to unseen sites. Because this study is diagnostic, all subsequent EG and memory analyses use the time-based split. The spatial strategies serve as robustness checks on the sensitivity of predictive skill to evaluation design, because adequate predictive accuracy is required before model-based attribution can be interpreted.

Methods Section 3.3 distinguishes these evaluation settings.

> "We explored multiple training and validation strategies as benchmarking analyses rather than treating these strategies as interchangeable approaches. The datasets were partitioned using four complementary strategies: time-based split, time-based cross-validation (time-split CV), site-based cross-validation (site-split CV), and leave-one-site-out (LOSO). The time-based split and time-split CV evaluate temporal generalization within observed sites, whereas site-split CV and LOSO evaluate spatial transfer to unseen sites."

Results Section 4.1 and Discussion Section 5.1 further state:

"Because our primary objective was to investigate EF dynamics and memory effects, we used the time-based split for the subsequent analyses, as it provided both strong predictive skill and a more suitable basis for an interpretable explainer model at each site."

"Model performance should therefore be interpreted in the context of both the target variable and the evaluation setting. Accordingly, the strongest performance obtained under the time-based split should be interpreted as temporal generalization within sites rather than as evidence of strong transferability to unseen sites. The lower performance under site-based cross-validation and LOSO is consistent with the broader ecosystem-flux literature showing that spatial generalization remains more challenging than within-site temporal prediction."

Comment 3:

Several aspects of the presentation of the memory effects also require clarification. It is not always clear exactly what the figures show, particularly Figures 8-10. Specifically, it should be stated explicitly in the figure captions and the main text whether the reported memory effect contributions are based solely on antecedent precipitation or whether other variables also contribute. This point appears in several places and should be clarified consistently throughout the manuscript.

Responses:

Thank you. We have clarified consistently in the Methods, Results, and captions that Figures 8, 9, and 10 use the full dynamic predictor set, S={P, Ta, RAD, VPD, WS, LAI}. These figures therefore show aggregated all-driver memory metrics, not precipitation-only contributions.

For instance:

"Figures 8–10 use the full dynamic predictor set."

"In addition to the memory decay timescales, for Figures 9 and 10 we quantified the memory effect using the cumulative antecedent contribution from lags 7 to 365 days relative to the total contribution from lags 0 to 365 days. In these analyses, we used the full dynamic predictor set, S={P, Ta, RAD, VPD, WS, LAI}."

All the related figure captions have been updated.

Comment 4:

The manuscript analyses more than 200 sites, yet only a small subset is presented in the main paper. The criteria for selecting these example sites are not explained. The authors should justify why these sites were chosen and discuss whether they are representative of the broader dataset.

Responses:

Thank you. Section 2.1 gives the complete sequence used to screen 215 sites and retain 90 sites, including variable availability, removal of managed cropland and wetland sites, preprocessing, and the minimum record-length criterion. The examples in Figures 4 and 5 were selected from these 90 sites to show clear drydown and attribution patterns across contrasting vegetation and hydroclimatic settings. They are illustrative rather than statistically representative; all quantitative conclusions use the full 90-site dataset. Please check our revised manuscript.

Comment 5:

The discussion of long-range influences of antecedent precipitation (e.g., around lines 660 and 699) is interesting but currently not well supported by the presented results. It is unclear where this effect is actually demonstrated. The authors should identify which sites exhibit such long-range influences and explicitly point readers to the corresponding analyses or figures.

Responses:

Thank you. We added Figure S6 and revised Section 5.2 to identify the sites and target dates. At DE-Obe (target date: 25 January 2009) and FR-Pue (target date: 30 November 2004), Expected Gradients show strong positive precipitation signals 364 and 363 days before the respective targets. Figure S6 presents the corresponding precipitation records and attribution profiles. The manuscript now states: "In our analysis, winter precipitation at DE-Obe and FR-Pue occurring close to 365 days earlier may continue to contribute to target-day EF predictions (Figure S6)."

Comment 6:

Finally, I think the manuscript would benefit from a stronger positioning within the existing literature. In particular, it would be valuable to explain more explicitly what new insights this study provides beyond previous work addressing similar questions with different methodologies, such as Teuling et al. (2006).

Responses:

Thank you. Teuling et al. (2006) characterized selected rain-free drydowns using a scalar ET decay timescale. In contrast, our LSTM-based explainable machine-learning framework reconstructs daily EF and provides prediction-specific, driver-resolved and lag-resolved attribution profiles across 90 sites. This retains day-to-day and event-to-event heterogeneity rather than reducing each drydown to a single timescale. This event-resolved capability is demonstrated in Figures 4 and 5.

> "This approach reveals the specific contributions of inputs to daily EF predictions." The Results also state: "Expected Gradients reveal the contributions of antecedent climate variables, including precipitation, air temperature, vapor pressure deficit (VPD), net radiation, and LAI, to current EF predictions."

These daily attribution profiles also provide a basis for future clustering of hydroclimatic regimes, including dry and wet seasons, extreme evaporative episodes, and periods of extreme air temperature or VPD. Such extensions could diagnose regime-dependent controls and generate mechanistic hypotheses (Sections 1, 3.5, 4.1, and 5.2).

Comment 7:

The discussion section could also be streamlined. Several arguments are repeated, particularly the interpretation linking rooting depth to the diagnosed memory effects. Reducing these repetitions would improve the readability and sharpen the overall discussion.

Responses:

Thank you. We removed repetitive statements in discussion section.

Reviewer: 3

General evaluation:

Overall, I think the authors have made extensive improvements to the manuscript. The framing of the contributions in the introduction is clearer, the description of the data sources, data processing and training procedure is now complete, and the presentation of the results and discussion better supports the goals of the study.

Responses:

We thank the referee for the positive assessment.

Comment 1:

There is one remaining issue that I recommend resolving. Figure 8 has been replaced in the revised manuscript, but the text appears to still address the old version (eg. lines 475-485). The methods section also describes precipitation-specific analyses. Please restore the previous figure 8 or revise the text accordingly.

Responses:

Thank you. We retained the revised Figure 8 and updated the corresponding text throughout the manuscript for consistency.