Page, J. C., Kauwe, M. G. D., Abramowitz, G., & Pitman, A. J. (2023). Non-Stationary Lags and Legacies in Ecosystem Flux Response to Antecedent Rainfall. *Journal of Geophysical Research: Biogeosciences*, *128*(5), [e2022JG007144]. https://doi.org/10.1029/2022JG007144

University of Bristol - Explore Bristol Research
General rights

Key Points:

- We use a *k*-means clustering plus regression approach to explore the time-varying response of terrestrial fluxes to antecedent climate
- The role of antecedent climate in ecosystem functioning is highly site- and time-dependent
- Site vegetation classification is a greater predictor for the precipitation memory of terrestrial fluxes than site aridity

Non-Stationary Lags and Legacies in Ecosystem Flux Response to Antecedent Rainfall

Jon Cranko Page[1,2] , **Martin G. De Kauwe**[3] , **Gab Abramowitz**[1,2] , and **Andy J. Pitman**[1,2]

[1]ARC Centre of Excellence for Climate Extremes, Sydney, NSW, Australia, [2]Climate Change Research Centre, University of New South Wales, Sydney, NSW, Australia, [3]School of Biological Sciences, University of Bristol, Bristol, UK

Abstract Ecosystem function can be affected directly by climate, including by meteorological extremes, and also by sustained lags and legacies on timescales that surpass those of the weather events themselves. However, important gaps remain in our understanding of the influence and timescale of persistence of antecedent climate, known as environmental memory, on terrestrial carbon and water fluxes. Identifying the interactions between the lagged response to climate and the legacies to climate extremes, and whether the influence of memory varies through time, has not been fully explored. Here, we used a novel *k*-means clustering plus regression approach to examine timeseries of the sensitivity of terrestrial fluxes to antecedent precipitation at 65 eddy-covariance sites across a range of ecosystems. Quantifying the sensitivity to past precipitation and temperature reveals that the role of memory in ecosystem fluxes varies across sites and in time. When memory was accounted for in the model, relative improvement in modeled site flux r^2 compared to an instantaneous model varied between 0% and 57%, with mean of 12%. Our results show that vegetation type was a stronger predictor of memory importance than site aridity, implying a need to understand vegetation resilience conferred by physiological traits and acclimation capacity. The influence of memory varied strongly through time at many sites, with the role of different timescales exhibiting consistent non-stationarity. Our results demonstrate the importance of accounting for time-varying vegetation response to antecedent rainfall in land surface models to accurately predict future terrestrial fluxes.

Plain Language Summary To predict how changes in future climate and weather extremes might impact terrestrial ecosystems, we need to understand the timescales of vegetation response to antecedent climate. Prevailing methods of exploration assume such responses to be stationary, that is constant through time. We present a novel approach that shows how the memory of plants to climate conditions change through time. We show that the carbon and water fluxes of vegetation can be significantly sensitive to antecedent rainfall and importantly that this sensitivity can vary substantially through time. Plant functional type is a key indicator of the role of memory to precipitation, while the response to antecedent rainfall is not determined by site aridity. Predicting future changes in the global carbon sink requires understanding how vegetation responds to climate across timescales. Identifying these timescales at which plants respond to climate is critically important as the climate changes, especially if extremes (e.g., heatwaves) become more frequent due to compounding effects.

1. Introduction

Climate change is increasing the frequency and intensity of some meteorological extremes, with implications for the terrestrial carbon and water cycles (Allen et al., 2015; Dunn et al., 2020; IPCC, 2021; Reichstein et al., 2013). Extreme rainfall has become more intense on short timescales (Allan & Soden, 2008; Min et al., 2011; X. Zhang et al., 2013), summer heatwaves have become more frequent and intense (Alexander, 2016; Perkins-Kirkpatrick & Lewis, 2020; Schär et al., 2004; Stott et al., 2004) and incidents of record-breaking and multi-year droughts have been observed globally (De Kauwe et al., 2022; Jiménez-Muñoz et al., 2016; Szejner et al., 2020; Williams et al., 2022). These weather extremes affect the vegetation through reductions in function (Ciais et al., 2005; Frank et al., 2015; Ma et al., 2016; Moran et al., 2014; Zscheischler et al., 2014a, 2014b) and in extreme cases, lead directly to mortality (Anderegg et al., 2015a; Arend et al., 2021). Understanding the vegetation's legacy to meteorological extremes, including the timescales of recovery, as well as the spatial recovery patterns, is emerging as a critical knowledge gap that limits our predictive capacity.

An increasing body of literature has investigated the timescales of memory that can affect vegetation function (e.g., De Boeck et al., 2018; Gong et al., 2020; Huxman et al., 2004; Mantoan et al., 2020; Peltier et al., 2018; Ogle et al., 2015; Schwalm et al., 2017; Vicente-Serrano et al., 2010; X. Wu et al., 2018). Here we differentiate the response of ecosystem-scale fluxes to antecedent climate based on both the type of the antecedent climate event and the response itself. We define a persistent but diminishing effect resulting from a climate extreme (or disturbance) as a "legacy". By contrast, a "lag" is a systematic consistent behavioral delay in response to the typical climate at any given time, acting at timescales from hours to a year or more. Both mechanisms are classed as "memory" effects for the influence of antecedent climate on current terrestrial ecosystem functioning.

Many studies have investigated the influence of seasonal (intra-annual) lags on ecosystem functioning. These often focus on the role of antecedent precipitation on carbon fluxes and growth. For instance, intra-annual rainfall patterns can be of greater importance to grassland biomass than total annual rainfall (Hovenden et al., 2014, 2018, 2019) while the same amount of precipitation received in fewer events can reduce terrestrial carbon uptake (Arca et al., 2021). Precipitation deficits experienced at different times of the year have varying impacts on future growth (Alves et al., 2020; Hahn et al., 2021; Huang et al., 2018, 2021), while some plants can acclimate to multiple droughts within the same growing season (Lemoine et al., 2018). Soil respiration, a component of the carbon flux, can have delayed responses to precipitation of around 2 days to 10 weeks (Cable et al., 2013; Cleverly et al., 2013). Carbon fluxes can be impacted by a rainfall event for anywhere between 5 days to 6 weeks in semi-arid woodlands (Cleverly et al., 2016), or up to 8 months in grasslands (T. Zhang et al., 2015). Other studies have revealed intra-annual lags between plant water content (affecting both carbon and water fluxes) and precipitation (Feldman et al., 2020), and carbon fluxes and temperature (T. Zhang et al., 2015). Overall therefore, there is ample evidence that climate, especially precipitation and temperature, at timescales of up to 1 year can influence current ecosystem carbon and water fluxes. Indeed, the prior year's climate can be a significant driver of current year productivity in some ecosystems (L. Liu et al., 2018; Sala et al., 2012; Ukkola et al., 2021; T. Zhang et al., 2015), indicating a role of inter-annual lags in ecosystem fluxes.

Legacies, in contrast to lags, are often identified over longer timescales. For example, precipitation deficits can impact tree ring growth for up to 5 years (Anderegg et al., 2015b; Vanoni et al., 2016), and sub-alpine forests experienced up to 11 years of legacy in mortality rates following droughts (Bigler et al., 2007). Legacy studies often consider the impact of large disturbances, such as fire impacting carbon fluxes for over 4 years (Sun et al., 2020) or consider ecosystem functions other than carbon or water fluxes, such as plant biomass and diversity being impacted by soil disturbances for up to 15 years into the future (Seabloom et al., 2020). A number of studies have examined whether a specific event produced legacy effects (Bastos et al., 2020; Griffin-Nolan et al., 2018; Peltier et al., 2021) which, while important, may miss compounding lags, confounding variables, or other events that influence the legacy effect (Bastos et al., 2021). Where studies consider legacies associated with multiple extreme events, these are often focused on growth metrics such as tree rings due to the requirement for long datasets that capture multiple events (Anderegg et al., 2015b; Vanoni et al., 2016). However, legacy effects seen in tree rings are not necessarily present in other aspects of ecosystem functioning, such as fluxes or leaf area (Kannenberg et al., 2019, 2020, 2022) and can be hard to directly connect to underlying processes represented in models. Some studies have instead focused on systematic lags that do not vary through time, using methods such as Stochastic Antecedent Modeling (Cranko Page et al., 2022; Guo et al., 2020; Ogle et al., 2015; Ryan et al., 2015, 2017) and lagged correlations (T. Zhang et al., 2015). These studies help our understanding of ecosystem functioning and are valuable in developing better LSMs (Keenan et al., 2012). However, as the rate of climate change intensifies, it is possible that systematic lags are themselves changing through time, which may have important implications for ecosystem function (Peltier & Ogle, 2020). Such changes, whether in the magnitude or timescale of influence of lags, would be missed under the assumption of memory stationarity (Peltier & Ogle, 2020).

Here, we introduce a novel method to explore the memory effects in ecosystem fluxes, by utilizing a machine learning approach to investigate which lags contribute the most to flux predictability at different timescales. This method enables information on both lags and legacies at a site to be extracted without any assumptions of behavioral stationarity, which are often inherent in prior studies (e.g., Cranko Page et al., 2022; Hovenden et al., 2014; L. Liu et al., 2018; Peltier et al., 2018; Richard et al., 2008). To demonstrate the power of this approach, we use it here to understand the role of memory to past precipitation and temperature in both net ecosystem productivity (NEP, the carbon flux) and latent heat (LE, the water flux). We compare metrics of memory importance with site characteristics such as aridity (Y. Liu et al., 2019) and plant functional types to investigate which site traits influence the role of lags and legacies in their ecosystem fluxes.

We examined the following hypotheses: (a) the inclusion of ecosystem memory improves the accuracy of modeled fluxes; (b) the sensitivity of fluxes to antecedent climate is non-stationary and varies both intra- and inter-annually; (c) ecosystem memory is more influential in the modeling of NEP than LE due to NEP having more mechanisms for delayed responses to prior conditions (i.e., via carbon losses from belowground respiration fluxes); and (d) the more arid a site is, the greater improvement in model performance when antecedent climate is included.

2. Methods

2.1. Data

2.1.1. Flux Data

Half-hourly meteorological and flux data from 65 eddy-covariance flux towers were taken from the PLUMBER2 forcing and evaluation datasets (Ukkola et al., 2022). This data set gap-fills and quality-checks the FLUX-NET2015 data set and was further aggregated to daily data for this study. Table S1 in Supporting Information S1 shows the list of sites included in this study. The sites are primarily located in North America (21 sites), Europe (33 sites), and Australia (9 sites), with one site apiece in French Guiana and Russia. The sites cover a range of vegetation types representative of the distribution amongst all FLUXNET sites, with 33 forest sites, 12 grasslands, and 10 cropland sites comprising the majority of the sites. The remaining sites are split into five savannah sites, two shrubland sites, and two wetland sites.

In this analysis, we examine LE and NEP fluxes. NEP, as a direct measurement of net ecosystem carbon flux, was preferred over flux-derived GPP due to known issues with partitioning between uptake and respiration (Renchon et al., 2021). The in situ meteorological data, namely total downward shortwave radiation (SW_{down}), mean air temperature (T_{air}), mean vapor pressure deficit (VPD), mean wind speed (W_S) and total precipitation (PPT), were used as predictors for the two fluxes in the machine learning setup described below.

2.1.2. Satellite Data

Leaf area index (LAI) data from the Copernicus Global Land Service was also used as a predictor for the fluxes in the same vein as the meteorological variables. The inclusion of LAI is intended to capture any differences in ecosystem functioning between growing and non-growing seasons. Note that this LAI data is included in the PLUMBER2 data sets, and is interpolated to daily data from the original temporal resolution (Ukkola et al., 2022). Full details of the LAI processing can be found in Ukkola et al. (2022).

2.1.3. Additional Data

To explore the relationship between memory effects and site aridity, two indices of site behavior were calculated so that site aridity could be categorized. These indices were the site Evaporative Index (EI), calculated as mean annual evapotranspiration (MAET) divided by mean annual precipitation (MAP), and Aridity Index (AI), calculated as mean annual potential evapotranspiration (MAPET) divided by MAP. Monthly potential evapotranspiration was calculated using Thornthwaite's method (Thornthwaite, 1948) and then aggregated annually. These indices were calculated from the PLUMBER2 data and therefore might differ slightly from the long-term site records. To explore the role that vegetation type plays in the expression of memory, International Geosphere-Biosphere Programme (IGBP) classifications of plant functional type (PFT) for each site were taken from the FLUXNET website (FLUXNET, 2023). Details of each class can be found in Strahler et al. (1999) and the class of each site is specified in Table S1 in Supporting Information S1.

2.2. Modeling

We model daily NEP and LE using an in-sample *k*-means clustering plus regression machine learning methodology (Jain et al., 1999; MacQueen, 1967). This approach models a target variable (here, each of the fluxes) using a set of predictor variables (the meteorological variables and LAI data), much like any model of this system would, except the relationship between flux and predictors is constructed entirely using observed site data without any physical assumptions. Hence, this is a purely empirical approach. *k*-means clustering partitions the observations (here, individual daily timesteps) of the predictors into groups, or clusters, by utilizing Thiessen polygons so that each cluster contains daily time steps with similar predictor values. For each cluster a multiple linear regression

Table 1
The Inputs to Each Model Implementation

Predictors	Model	Categorization
LAI(t)	Instantaneous	–
T_{air}(t)	Instantaneous	–
SW_{down}(t)	Instantaneous	–
VPD(t)	Instantaneous	–
W_S(t)	Instantaneous	–
PPT(t)	Instantaneous	–
T_{air}(t-1)	Historical	–
T_{air}(t-2) to T_{air}(t-7)	Historical	–
T_{air}(t-8) to T_{air}(t-14)	Historical	–
T_{air}(t-15) to T_{air}(t-30)	Historical	–
PPT(t-1) to PPT(t-30)	Historical	Seasonal
PPT(t-31) to PPT(t-90)	Historical	Seasonal
PPT(t-91) to PPT(t-180)	Historical	Seasonal
PPT(t-181) to PPT(t-365)	Historical	Mid-term
PPT(t-366) to PPT(t-730)	Historical	Mid-term
PPT(t-730) to PPT(t-1095)	Historical	Long-term
PPT(t-1096) to PPT(t-1460)	Historical	Long-term

Note. t refers to the current day. For historical T_{air} and PPT, the sum of the daily data is calculated over the period covered. The categorization column refers to how the historical precipitation predictors are grouped into seasonal, mid-, and long-term "memory."

is performed between the predictors and the fluxes for the daily time steps that belong to the cluster, as though each time step were entirely independent.

For each site and each flux, two separate models were developed. The first model, referred to as the instantaneous model, used only current-day meteorological and LAI data as predictors for the NEP and LE fluxes. The second implementation, referred to as the historical model, included lagged rainfall data from up to 4 years in the past and lagged temperature from the prior month to the instantaneous model. The chosen temporal periods for the lagged predictors are based on both prior studies (Cranko Page et al., 2022; Y. Liu et al., 2019) and physical mechanisms, as rainfall has clear storage pools (namely soil moisture and groundwater) while temperature storage (e.g., soil temperature) is less likely to influence fluxes on longer timescales. Table 1 shows the complete list of predictors. For each model implementation, the entire predictor set was included in both the clustering and the linear regression calculations.

To minimize overfitting, any clusters that contained fewer observations than 10 times the number of predictors (i.e., any cluster containing fewer than 60 observations for the instantaneous model, or fewer than 170 observations for the historical model) were expanded during the regression phase. Based on Euclidean distance from the cluster center, the closest observations not already in the cluster were included in the linear regression such that the regression included 10 times the number of predictors. Only the observations initially assigned to the cluster were then assigned the modeled flux values; the expansion observations were still assigned the modeled flux value from the cluster they were assigned during the clustering algorithm.

The clustering was performed with 729 clusters for four reasons (cf. Cranko Page et al. (2022), which used between two and eight clusters). First, a greater number of clusters improves the model performance and, due to the regression expansion mechanism described above, did not risk overfitting. Second, we prefer a number of clusters that captures something of the underlying physical behavior in the data set. With six predictors in the instantaneous model, if we assume that each predictor might have clusters where its values are "low," "average," and "high," all possible combinations of these three values across the predictors will result in 729 clusters (3^6). To ensure a fair comparison, we kept the same number of clusters for the historical model. Additionally, we wanted at least two years of modeled fluxes to explore, thus 730 clusters was the maximum number of clusters possible since each day is treated as a single observation, and having more clusters than the total number of observations will inevitably result in empty clusters. Sensitivity tests on the number of clusters showed that, as the number of clusters increased, the results tended toward stable solutions as the clusters increasingly needed to bring in data from neighboring clusters to complete the regression stage (not shown). Finally, the greater cluster resolution of this framework compared to prior studies allows a much more nuanced description of the evolution of memory responses to antecedent climate through time, providing insight into which variables and memory timescales provide predictability at a particular point in time at a given site.

All flux, meteorological and LAI data were scaled on a site-by-site basis such that the full site time series had mean of zero and standard deviation of one. While necessary for the *k*-means clustering to ensure predictors of larger scales do not dominate the clustering, this scaling has the additional benefit of allowing all results to be directly comparable across sites, models, and predictors.

2.3. Analysis

Multiple linear regressions were fitted separately for each cluster in a given site's model, with all predictors normalized beforehand (Equation 1). The coefficient of each predictor, given that all predictors had been normalized, is representative of the sensitivity of the flux to that predictor within the cluster. We can therefore examine flux sensitivity to each predictor through time by taking the value of this coefficient at each timestep based on the cluster that each timestep belongs.

$$\text{Flux}(t) = \text{Intercept} + \sum_i a_i \times \text{Predictor}_i(t) \tag{1}$$

Next, regression coefficients were masked based on two criteria to ensure that only well-defined sensitivity was reported. For each predictor, timesteps in the coefficient timeseries where any of the masking criteria were true had the coefficient value replaced with "NaN" values. First, we masked any timesteps where the regression coefficient in the linear regression was not significant at a p-value of 0.05 or less. Since these coefficients are statistically non-significant, it would be erroneous to report any subsequent behavior attributed to them as realistic. Additionally, any clusters where the linear regression between the observed and modeled flux had an r^2 value of less than 0.2 were also masked. Although this choice of a low r^2 is arbitrary, the chosen value is a reasonable limit at which to determine that the linear regression captures a reasonable degree of the variability. This ensures that sensitivities resulting from poor model performance are not conflated with more significant findings in our results. These "NaN" masks were only applied after metrics of model accuracy were calculated (see Figures 1 and 2).

2.4. Model Implementation

All analysis was performed using R software (R Core Team, 2020). The k-means clustering was performed using the "KMeans_Rcpp" function from the "ClusterR" package (Mouselimis, 2022). The linear regressions were fitted using the "lm" function. PET was calculated using the "thornthwaite" function from the "SPEI" R package (Beguería et al., 2014; Vicente-Serrano et al., 2010).

The k-means clustering was implemented using the k-means++ initialization (Arthur & Vassilvitskii, 2007). This method aims to produce initial clusters that cover as much of the occupied sample space as possible, rather than choosing the initial centers at random as per the standard k-means algorithm. This initialization technique has been shown to improve both the accuracy and speed of the k-means clustering (Arthur & Vassilvitskii, 2007).

3. Results

3.1. Model Predictions at Eddy-Covariance Flux Sites

We first demonstrate the high-fidelity in the machine learning approach to simulate timeseries of ecosystem fluxes, which builds confidence in our interpretation of regression coefficients and the sensitivity of fluxes to predictors. Figure 1 shows observed and modeled LE and NEP for two sites each, using both the instantaneous and historical models. Figure 1 demonstrates the ability of the historical model to accurately model the observed flux, and the extent to which the historical model can outperform the instantaneous model (see Table 2 which summarizes model performance metrics for the instantaneous and historical models for both fluxes as well as Figure S66 in Supporting Information S1). For example, Figure 1b shows the predicted LE flux in 2011–2012 years is higher in the summer using the historical model than the instantaneous model, in line with the observations. This improved prediction skill appears to carry forward into the 2012–2013 predictions, where the historical model and observations show a reduced summer flux, compared to the instantaneous model, in 2011–2012.

For both fluxes at both daily and monthly timesteps, the historical model performs better than the instantaneous model in all metrics apart from the mean bias error. Similar results are shown in Figures S1–S65 in Supporting Information S1 for all included sites. Several studies have shown that similar approaches routinely outperform physically based ecosystem models or LSMs, especially when tested in-sample, as we are doing here (Abramowitz et al., 2008; Best et al., 2015; Nearing et al., 2018).

3.2. Examining the Importance of Memory Across Sites

Figure 2 shows the overall relative improvement in two model performance metrics (normalized mean error and r^2) between the instantaneous model and the historical model, highlighting the degree to which memory of precipitation and temperature affects site predictability. Results are shown on a Budyko curve (Budyko, 1974) to examine whether water or energy limitation plays a role in the importance of memory effects relating to precipitation and temperature. For both fluxes and both metrics, the historical model (i.e., accounting for past meteorological conditions) performs better at all sites than the instantaneous model. Relative improvement in normalized mean error (NME) is similar across both fluxes with a mean improvement of 24% for NEP and 23%
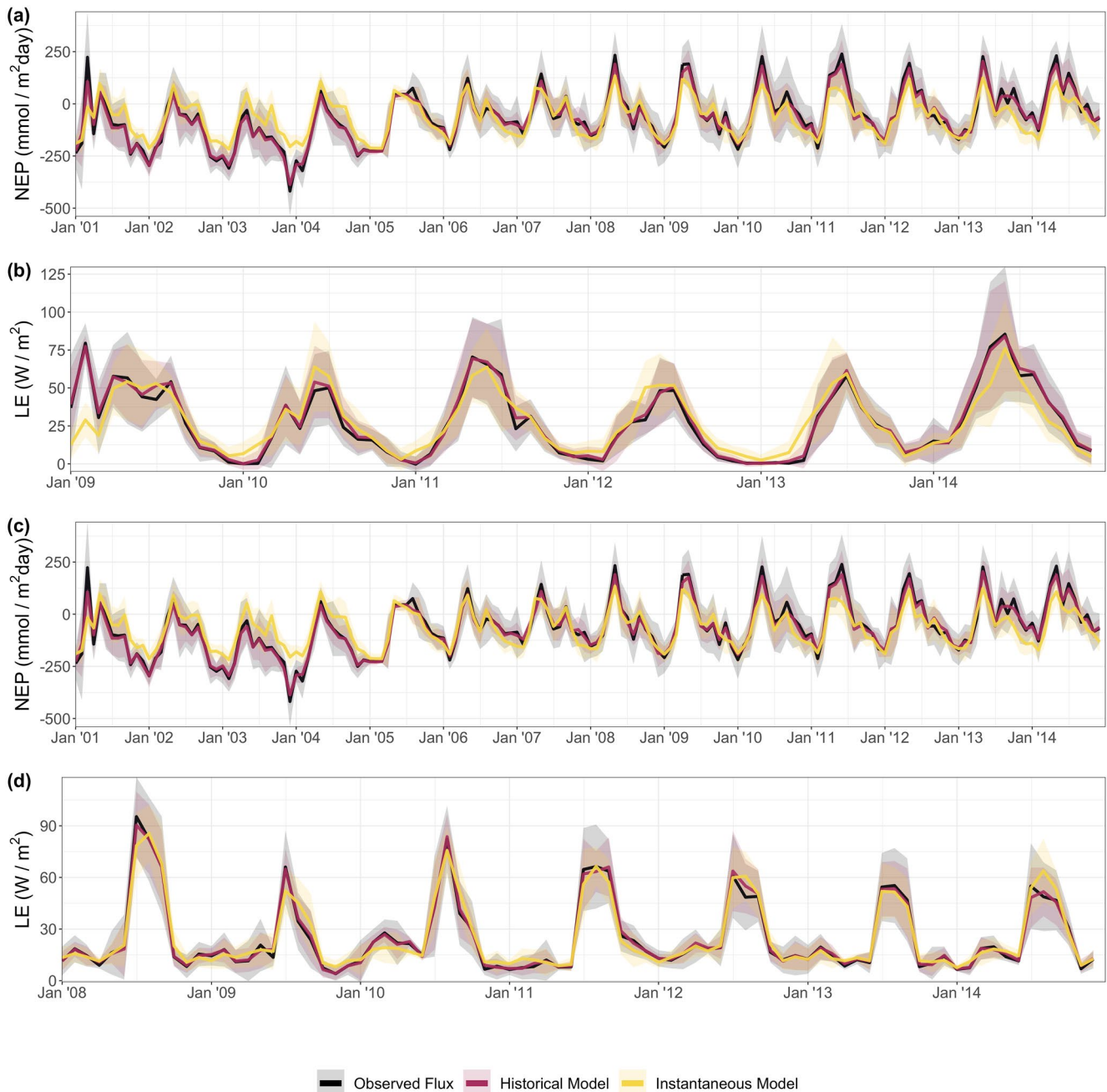
Figure 1. Timeseries of observed flux against the flux modeled using the instantaneous and historical models. The *x*-axis is the date, the *y*-axis is the flux value, the lines are the monthly mean daily flux and the shaded areas indicate the monthly standard deviation of the daily flux. Subplot (a) is the net ecosystem productivity (NEP) flux at CH-Dav, an evergreen needleleaf forest site, (b) is the LE flux at DE-Kli, a cropland site, (c) is the NEP flux at DK-ZaH, a grassland, and (d) is the LE flux at US-SRM, a savannah site. These four sites were chosen to illustrate the potential of our approach and are sites with some of the best model improvement when accounting for the memory effect.

for LE. Relative improvement in r^2 is smaller than for NME, especially when considering LE. For NEP, mean r^2 improvement is 17% while for LE the mean relative improvement is 7%. However, the mean r^2 for the instantaneous LE model is 0.86 which results in an upper limit on potential mean relative improvement of 16% (compared with a mean instantaneous r^2 of 0.73, and hence a limit of 37%, for NEP).

Relative improvement in both metrics and across both fluxes was not correlated with site aridity or the evaporative index, as shown by the r^2 and *p*-values in the lower right of the plots (Figures S67 and S68 in Supporting Information S1). This can be seen in Figure 2, where the relative memory improvement for a site is clearly not
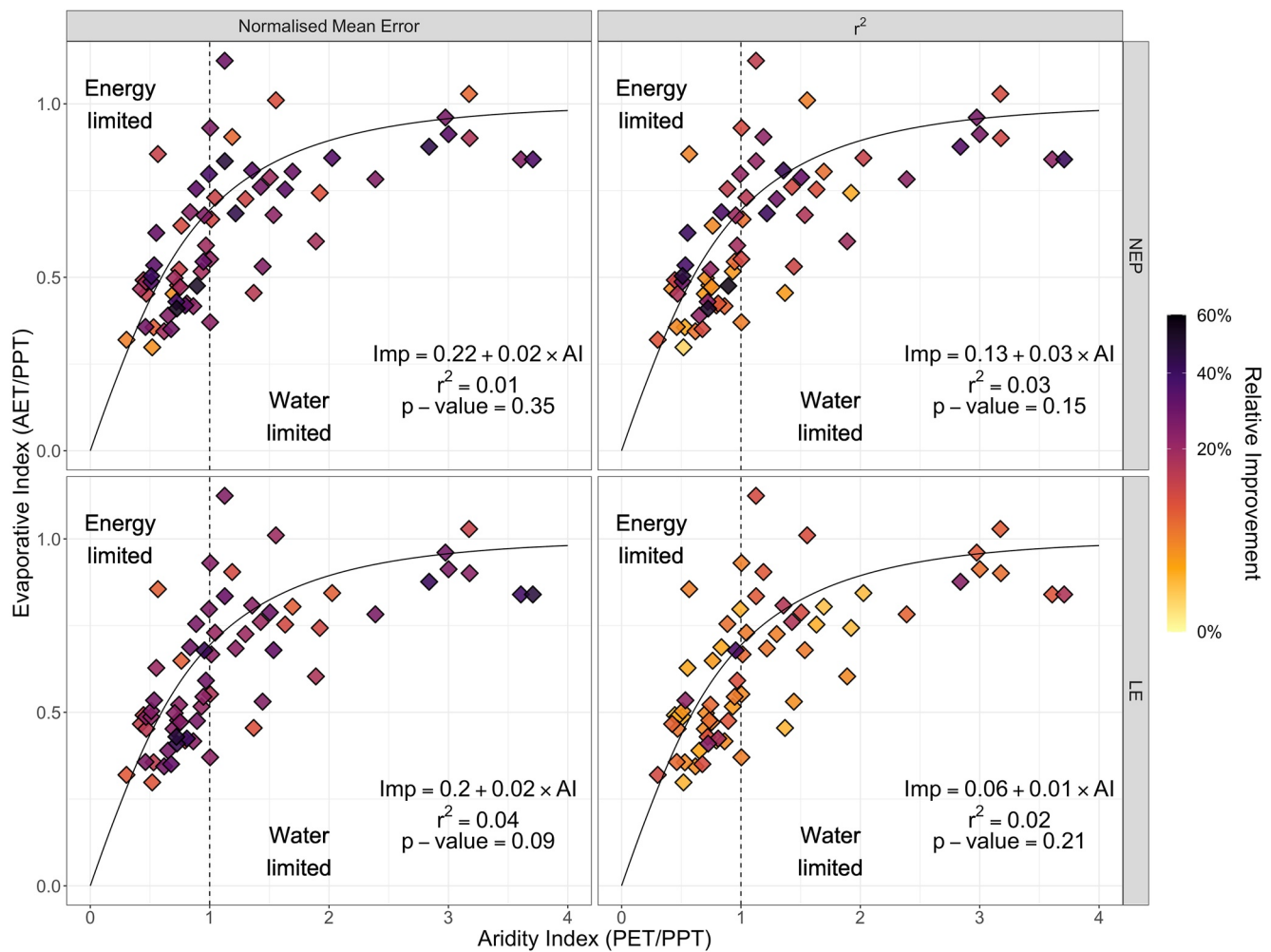
Figure 2. Relative improvement in model performance metrics between the instantaneous and historical models for each site. The plots are split by flux (net ecosystem productivity (NEP) top and LE bottom) and metric (NME left and r^2 right). The x-axis is the site aridity index of potential evapotranspiration over precipitation (AI = PET/PPT) and the y-axis is the site evaporative index of actual evapotranspiration over precipitation (EI = AET/PPT). The color of each point indicates the relative improvement in the metric value between the instantaneous and historical models (e.g., if the instantaneous model has $r^2 = 0.5$ and the historical model has $r^2 = 0.75$, then the relative improvement will be 50%). The solid black line indicates the Budyko curve, and the dotted black line indicates AI = 1, with sites where AI < 1 being considered "energy limited" and for AI > 1 "water limited." Correlation statistics are provided in the lower right of each panel, where "Imp" refers to the relative improvement.

related to the site's aridity index (x-axis), evaporative index (y-axis) or position along the Budyko curve. Similarly the distance of a site from the Budyko curve, indicating the presence of water sources/sinks outside of precipitation and evapotranspiration, also has no effect on the role of memory at the site. Relative improvement was also not correlated with the coefficient of variation of yearly precipitation, a proxy for site variability (Figures S69 in Supporting Information S1).

Notably, the seven sites with the greatest relative improvement in r^2 for NEP, all with an improvement of 40% or greater, are grasslands or croplands. There are some sites where environmental memory to precipitation and/or temperature plays a large role in both fluxes. For instance, DE-Kli, a cropland in Germany, had the second greatest improvement in r^2 amongst all sites for both NEP (51%) and LE (24%). Other sites where both fluxes have a large influence from antecedent climate include AU-Ync (41% improvement for NEP, 17% for LE), DK-ZaH (35% for NEP, 24% for LE) and CH-Dav (38% for NEP, 22% for LE). Alternatively, there are sites where the role of memory is much more important in one flux than another (relative to the other sites in this study). As an example, FR-Pue has a relative improvement in r^2 of 9% for NEP while LE has an r^2 improvement of 18%. Similarly, IT-SRo has only a 20% improvement in NEP r^2 but a 36% improvement for LE.

Table 2
Performance of Instantaneous Versus Historical Models

Timescale	Flux	Model	Aggregated year classifications (no. of sites)			Mean performance metrics			
			Poor mean	Poor variability	Acceptable	NME	MBE	Correlation	SD difference
Daily	NEP	Instantaneous	38	6	21	0.46	−0.007	0.85	0.20
		Historical	35	0	30	0.35	−0.019	0.92	0.14
	LE	Instantaneous	29	2	34	0.31	−0.002	0.93	0.10
		Historical	25	0	40	0.24	−0.004	0.96	0.07
Monthly	NEP	Instantaneous	12	38	15	0.35	−19,444	0.92	0.18
		Historical	3	0	62	0.15	−47,715	0.99	0.07
	LE	Instantaneous	2	36	27	0.20	−1,761	0.97	0.07
		Historical	2	0	63	0.08	−9,685	1.00	0.02

Note. "Daily" timescale refers to daily timeseries, "Monthly" refers to the daily timeseries aggregated to monthly by taking the mean. For the "No. of Sites" columns, the three categories applied to the aggregation of the timeseries into groups of each calendar day or month using means. For instance, the daily timeseries was aggregated by finding the mean flux for all 1 January, 2 January, and so on for all 365 days. The monthly timeseries was aggregated by taking the mean of all Januarys, all Februarys, and so on for all 12 months. A site has "Poor Mean" if any aggregated timestep is such that the modeled mean flux ± standard deviation falls outside the range of the observed mean flux ± standard deviation. A site has "Poor Variability" if it is not classified as having "Poor Mean" and at least 25% of timesteps (i.e., 90 days for "Daily" and 3 months for "Monthly") have a modeled standard deviation that is more than 50% larger or smaller than observed standard deviation. "Acceptable" sites are those that do not fall in either of the prior two classifications. The percentages sum to one. The performance metrics are the mean values for the 65 sites. At each site, the performance metrics are calculated for the entire timeseries. "NME" is the normalized mean error, "MBE" is the mean bias error, "Correlation" is the Pearson correlation coefficient, and "SD Difference" is the standard deviation difference.

For NEP, site improvement in r^2 was significantly correlated with improvement in NME ($r^2 = 0.67$, *p*-value <0.001). The improvement in the two metrics was also correlated for LE ($r^2 = 0.61$, *p*-value <0.001). For r^2, relative improvement for NEP was significantly correlated with improvement for LE ($r^2 = 0.38$, *p*-value <0.01). The same was true for NME ($r^2 = 0.51$, *p*-value <0.001). This implies that the importance of antecedent climate for a site's fluxes was robust between carbon and water fluxes, as well as being consistent between model performance metrics.

3.3. The Influence of Different Timescales of Ecosystem Memory

We next look at the role of different timescales (seasonal, mid-term, and long-term) of lagged rainfall on flux predictability (Figure 3) (see Table 1 for how rainfall predictors are categorized). Figure 3 illustrates the general "well-defined" median sensitivity of the site flux to each timescale of antecedent rainfall (seasonal, mid-term, and long-term); that is, 50% of significant timesteps have a sensitivity with magnitude equal to or greater than the sensitivity shown in Figure 3. The sensitivity of both NEP and LE to different timescales of antecedent rainfall was highly variable amongst sites. In general, the flux sensitivity to antecedent rainfall was greater for NEP than for LE. The mean of the median sensitivity across all categories was 0.55 for NEP and 0.40 for LE. For NEP, seasonal antecedent rainfall had a mean site median sensitivity of 0.50, increasing to 0.61 for mid-term antecedent rainfall before decreasing to 0.53 for long-term antecedent rainfall. Similarly, for LE, seasonal antecedent rainfall had a mean site median sensitivity at 0.40, mid-term median sensitivity is greatest at 0.44 and then the median sensitivity decreases to 0.37 for long-term antecedent rainfall. Overall, based on site median sensitivity, NEP was more sensitive to mid-term historic precipitation relative to other timescales while LE was influenced by the shorter timescales of seasonal and mid-term rainfall.

The Budyko curve in Figure 3 again shows how site aridity was not correlated with NEP median sensitivity at any timescale of historical precipitation. The standard deviation of NEP median sensitivities across timescales, that is the variance in sensitivity at different lag times, was also not correlated with site AI. However, for LE, seasonal median sensitivity was positively correlated with AI with the correlation weakening as the timescales lengthened. The standard deviation amongst LE median sensitivities for each timescale at each site was similarly
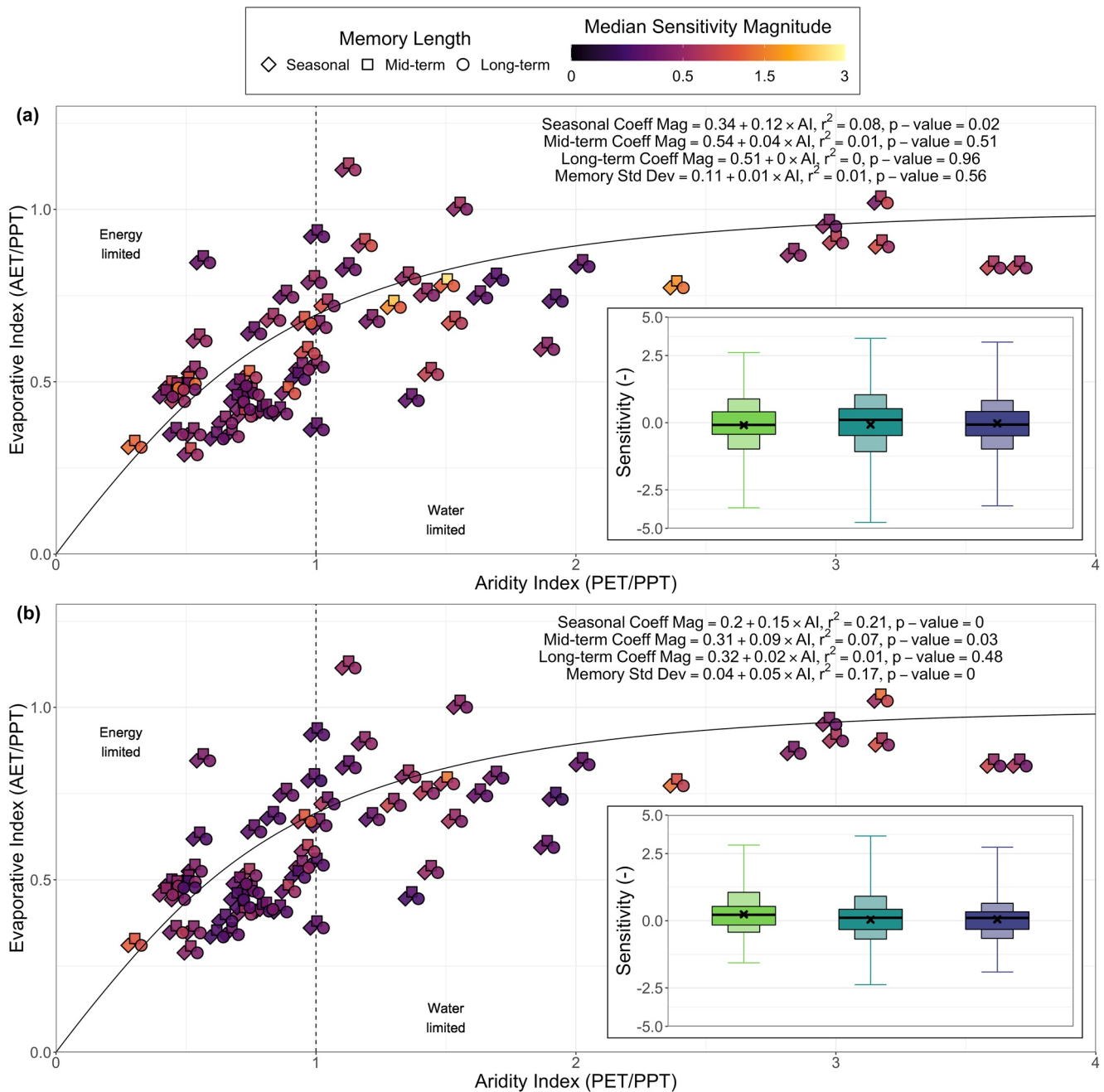
Figure 3. The median of absolute coefficient values from the historical model for each site and each category of historical rainfall for (a) net ecosystem productivity (NEP) and (b) latent heat (LE). The x-axis is the site aridity index of potential evapotranspiration over precipitation (AI = PET/PPT) and the y-axis is the site evaporative index of actual evapotranspiration over precipitation (EI = AET/PPT). Each site has three points jittered around its location in AI/EI space, one for each timescale of past rainfall. The color represents the value of the median of the absolute coefficient values for all predictors within the category and all timesteps. Lighter colors represent a higher coefficient magnitude and therefore a greater sensitivity. The inset plots shows box plots of the non-masked coefficients for seasonal, mid-term and long-term antecedent rainfall (left to right, respectively) for all sites and the corresponding flux. For each boxplot, the thick black line is the median, the wide boxes denote the interquartile range (IQR, 25th–75th percentile), the thin boxes mark the 10th to 90th percentile and the error bars denote the 1st and 99th percentiles. The black cross indicates the mean. Note the pseudo-log scale on the y-axis. Correlation statistics are provided in the upper right of each panel.

weakly correlated with site AI. This means that, as site aridity increased, the variation in sensitivity across timescales also increased.

Of note in Figure 3 are a selection of sites with marked sensitivity differences between seasonal, mid- and long-term antecedent rainfall. For NEP, one example was US-FPe, a grassland with AI 1.5 and EI 0.79, which had a

Table 3
r² Values of Pairwise-Correlation Between the Median Sensitivities for Each Site (n = 65) for Net Ecosystem Productivity (NEP) and Latent Heat (LE)

Precipitation timescale categories	NEP	LE
Seasonal × Mid-term	0.66	0.49
Seasonal × Long-term	0.46	0.40
Mid-term × Long-term	0.66	0.72

Note. All correlations were significant at a *p*-value <0.001.

median sensitivity to seasonal antecedent rainfall of 1.17, a mid-term median sensitivity of 2.60 and a decreased median of 1.15 for long-term timescales. AU-DaS, an Australian savanna site (with AI = 1.30 and EI = 0.73), similarly had a substantially larger sensitivity to mid-term antecedent rainfall (seasonal median sensitivity = 1.60, mid-term = 2.23, long-term = 1.14). Another grassland, CH-Oe1 (with AI of 0.51 and EI of 0.50) had a large increase in median sensitivity as the timescale becomes longer (seasonal = 0.43, mid-term = 1.00 and long-term = 1.30). Similarly, BE-Lon (cropland, AI = 0.89, EI = 0.48) had very strong median sensitivities to mid- and long-term antecedent rainfall (1.14 and 1.00 respectively) but weak seasonal sensitivity of 0.34. For LE, there were fewer sites with high variation amongst median sensitivities, although US-FPe continued to have a substantially stronger median sensitivity to mid-term antecedent rainfall compared to seasonal and long-term timescales (seasonal median = 0.63, mid-term = 1.41, long-term = 0.58).

For each timescale, the median NEP sensitivity was significantly correlated with median LE sensitivity ($r^2 = 0.75$, 0.50, and 0.52 for seasonal, mid- and long-term respectively, *p*-values < 0.001). This implies that the site sensitivity to antecedent rainfall was consistent between fluxes; that is, a site with high NEP sensitivity to antecedent rainfall at any timescale was likely to also have a (relatively) high LE sensitivity to the same timescale. Median sensitivities were positively correlated pairwise across the three timescale categories for both fluxes (Table 3). Hence, in general, a site that was sensitive to, for example, seasonal rainfall was also sensitive to both mid-term and long-term antecedent rainfall. Notably, the correlations between mid-term and long-term sensitivity are amongst the strongest which may suggest that the mechanisms resulting in sensitivity at the 6 months to 2 years timescale were similar to the mechanisms driving sensitivity to 2- to 4-year lagged antecedent rainfall. Correlations between seasonal and long-term memory sensitivities were the weakest, as might be expected due to the lack of continuity between the timescales.

Figure 4 illustrates the distributions of historical temperature and precipitation coefficient values for each timescale (temperature is only plotted at the monthly scale) and PFT grouping for the two fluxes. Positive NEP values imply an uptake of carbon and, therefore, an increase in sensitivity indicates that uptake increased and decreased as rainfall increased and decreased, respectively. It is important to note that, due to the varying sign of the sensitivity coefficients, a mean or median of zero in Figure 4 is not necessarily an indicator of low sensitivity. Instead consider that the thin boxes and error bars, exclusive of the larger box, indicate sensitivity magnitudes experienced at least 40% of the time (5th to 95th percentile, excluding the 25th–75th percentiles).

NEP in savannas, shrublands, and grasses was more sensitive to all timescales of antecedent precipitation compared to forests and wetlands (Figure 4a). This follows from both the greater magnitude of mean sensitivity, and the larger spread of the box plots, which indicates that they experience higher sensitivities. Interestingly, NEP within all PFT classes experienced periods of both negative and positive sensitivity to antecedent rainfall at all timescales. Negative sensitivity (i.e., an increase in rainfall leads to a decrease in carbon uptake and vice versa) was most common in savannas and grassland ecosystems. Savanna NEP had a particularly high sensitivity to rainfall at all timescales, although this switches from strong negative sensitivity at seasonal and mid-term timescales to a strong positive sensitivity at the long-term timescale. Shrubland NEP had a positive sensitivity to seasonal rainfall over 50% of the time (median in first panel of Figure 4a) while negative sensitivity to long-term rainfall is more frequent than positive. The inverse is seen in grasslands where most of the time NEP was negatively sensitive to seasonal rainfall but was more likely to be positively sensitive (increased NEP uptake with increased rainfall) to long-term rainfall. Wetland carbon uptake was positively influenced by increases in seasonal, mid-term and long-term rainfall, although sensitivity was generally of a smaller magnitude than at shrublands and grasslands. While the range of NEP sensitivity tends to decrease over time for shrubs/savannas and grasses, sensitivity at forest sites was mostly consistent across timescales. Forest PFT classes, except for evergreen broadleaf forests, did not have substantial changes in the range, median, or mean of NEP sensitivity to precipitation as the timescale extends. The greater variability for EBFs likely reflects the broader range of species/climate that this category encompasses (*i.e.*, eucalypts from Australia, tropical tree species, and *Quercus ilex*).

The relative consistency of forest flux sensitivity across precipitation timescales was also true for LE (Figure 4b). However, for deciduous broadleaf, evergreen broadleaf and evergreen needleleaf, seasonal sensitivity was
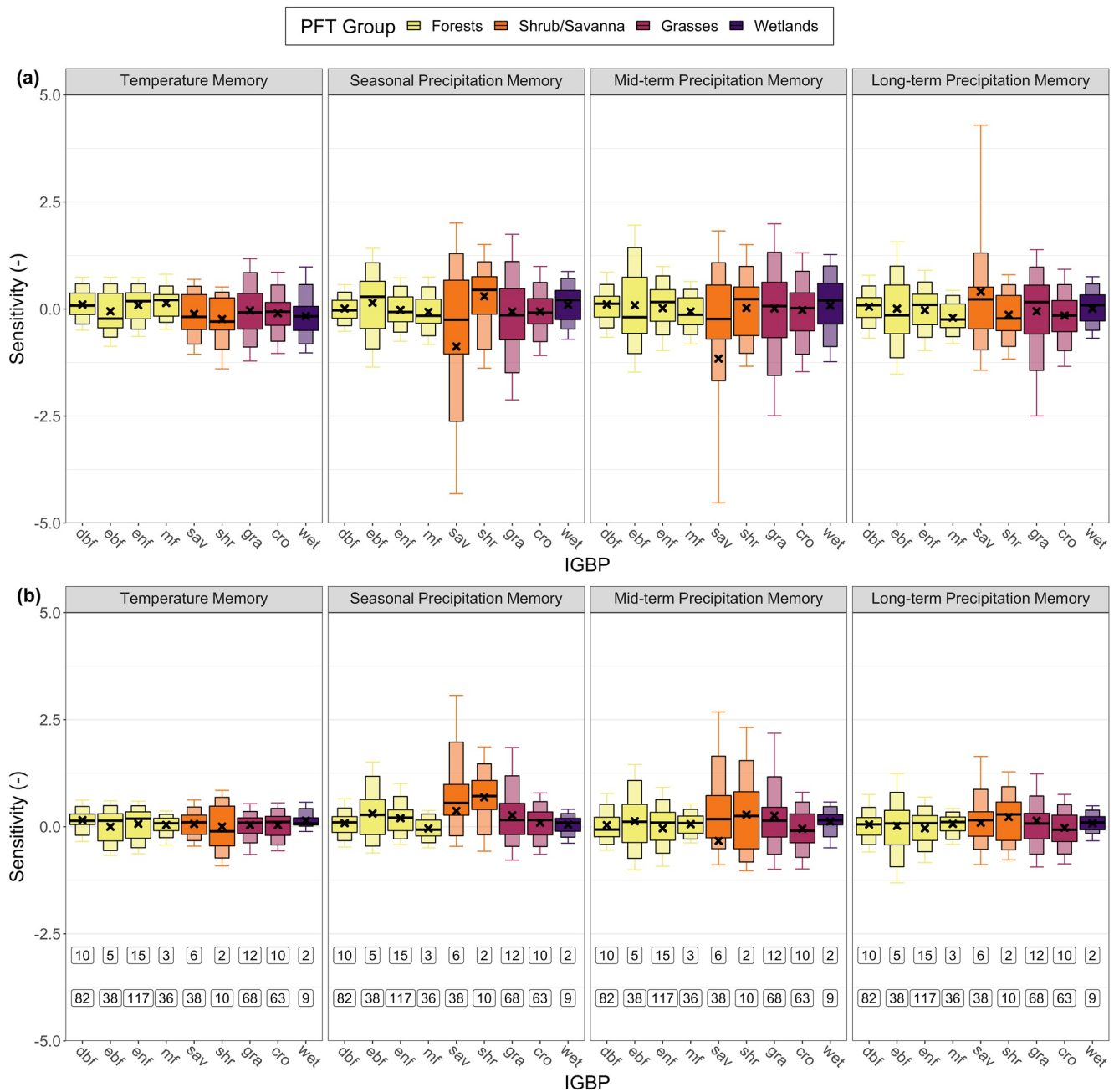
Figure 4. Boxplots of temperature and precipitation coefficient values grouped by International Geosphere-Biosphere Programme (IGBP) classification and split by historical rainfall category for (a) net ecosystem productivity (NEP) and (b) latent heat (LE). The *x*-axis is the IGBP classification (note that "osh" and "csh" were grouped in "shr" and "wsa" was combined with "sav"). The *y*-axis is the coefficient value (dimensionless due to the scaling of the model inputs). For each boxplot, the thick black line is the median, the wide boxes denote the interquartile range (IQR, 25th–75th percentile), the thin boxes mark the 10th to 90th percentile and the error bars denote the 5th and 95th percentiles. The black cross indicates the mean. The labels at the bottom of the plots are the number of sites (above) and the number of site-years (below) within the IGBP class. Outliers are not plotted due to a few instances of exceptionally large coefficients.

dominantly positive. As such, these forest biomes generally experienced an increase in evapotranspiration when there was an increase in seasonal rainfall. This dominant positive sensitivity was also apparent for the shrub/savanna and grassland sites. As antecedent rainfall timescales increased, LE sensitivity decreased and became more symmetrical with negative influence becoming as frequent as positive sensitivity. The lowest sensitivities across all PFT classes are seen for the LE fluxes to long-term antecedent rainfall. When both fluxes were considered together, shrublands, savannas, and grasslands experienced the greatest sensitivities to antecedent rainfall across all timescales.
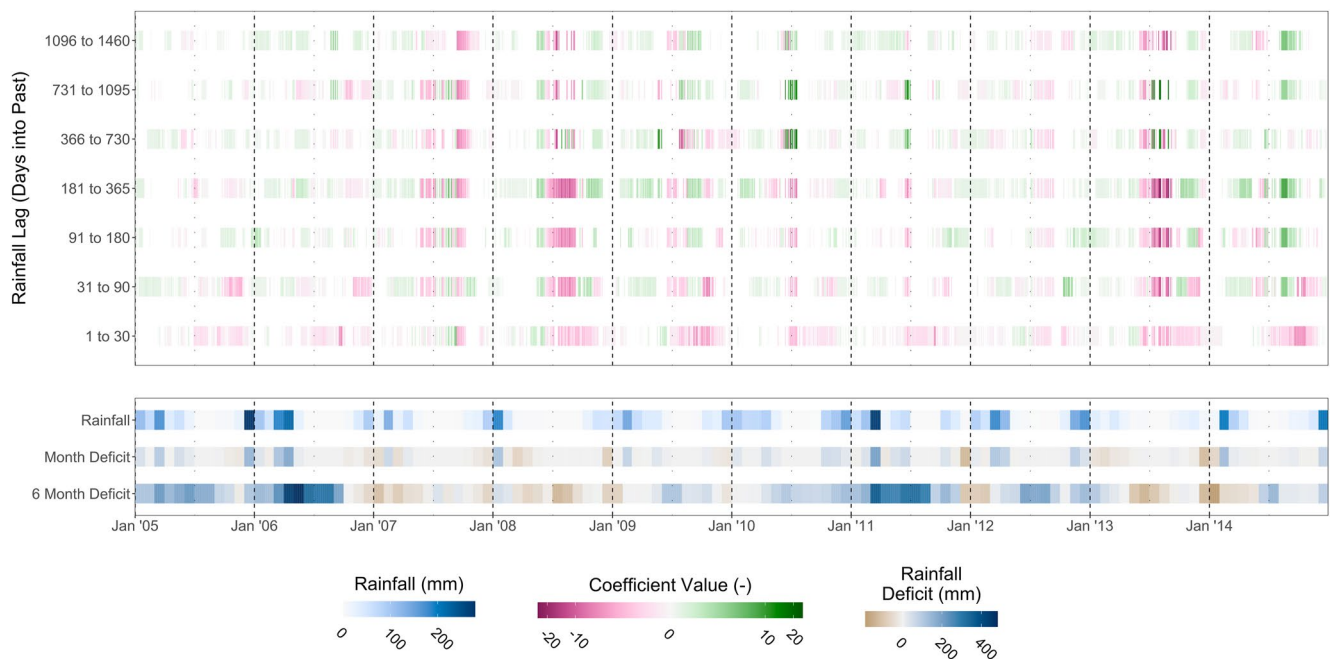
Figure 5. Timeseries of coefficient values for antecedent rainfall from the *k*-means clustering plus regression historical model for net ecosystem productivity (NEP) at US-Var. The *x*-axis is the date at a daily timestep, and the *y*-axis is each antecedent rainfall predictor, together with monthly rainfall totals (Rainfall), the difference between actual monthly rainfall and mean monthly rainfall for the calendar month (Month Deficit), and a rolling 6 month sum of the monthly deficit (6 Month Deficit). Each predictor is colored at each timestep according to the coefficient value calculated for the cluster in which the timestep belongs. Red timesteps indicate a negative sensitivity of NEP to rainfall (more antecedent rainfall within the period represented by the predictor results in less NEP) and green indicates a positive sensitivity (more antecedent rainfall results in greater NEP).

The flux sensitivity to antecedent temperature lagged up to 1 month is consistently low across PFTs and for both NEP and LE. Even for those PFT classes with relatively low sensitivity to antecedent precipitation, the sensitivity to antecedent temperature is only ever of a similar magnitude or smaller.

3.4. The Influence of Memory Is Not Stationary

The novel implementation of the machine learning approach also allows us to visualize the timescales of influence as a timeseries at each flux site, allowing an examination of whether the influence of memory changes as a function of time. Figures 5–8 show the timeseries of regression coefficients for each historical rainfall predictor at four different sites. These sites were chosen as they illustrate a selection of potential behaviors that can be identified in these types of plots. The same figures for the other 61 sites are included in the Supplementary Information. In Figure 5, the coefficient values for the NEP model at US-Var, a Californian grassland with a Mediterranean climate, are shown. The coefficients were greatest at times of low rainfall (approximately May to September for years 2007, 2008, and 2013 and to lesser extents in 2009, 2010, 2011, and 2014) and the greatest sensitivity was experienced during the driest year, 2013. Within the dry summers, rainfall up to one year into the past negatively affected NEP in all years but 2014. The effect of rainfall from one to four years into the past was generally weaker and varied between negative and positive sensitivity. Figure 6 shows the regression coefficients for NEP modeled at US-SRM, a woody savanna in southern Arizona. At this site, there was a clear positive memory effect of prior rainfall affecting NEP in late summer. The exact strength varies but timescales of between three to 6 months consistently exhibit the strongest influence. There are occasional strong negative sensitivities in the memory to precipitation. The seasonal timescales are consistently of greater influence than the mid- and long-term predictors. The coefficients for US-Ne3, a non-irrigated cropland in Nebraska, are shown in Figure 7. At this site, crops are present between approximately May to October, and the site experiences an extremely seasonal climate with cold winters and hot summers. The sensitivity of NEP to prior rainfall clearly reflects the period during which crops are present. However, the magnitude and direction of sensitivity varies between years. This can be compared to US-Ne1 and US-Ne2, two irrigated croplands, which have much more consistent sensitivity to antecedent rainfall (see Figures S54 and S55 in Supporting Information S1), suggesting a
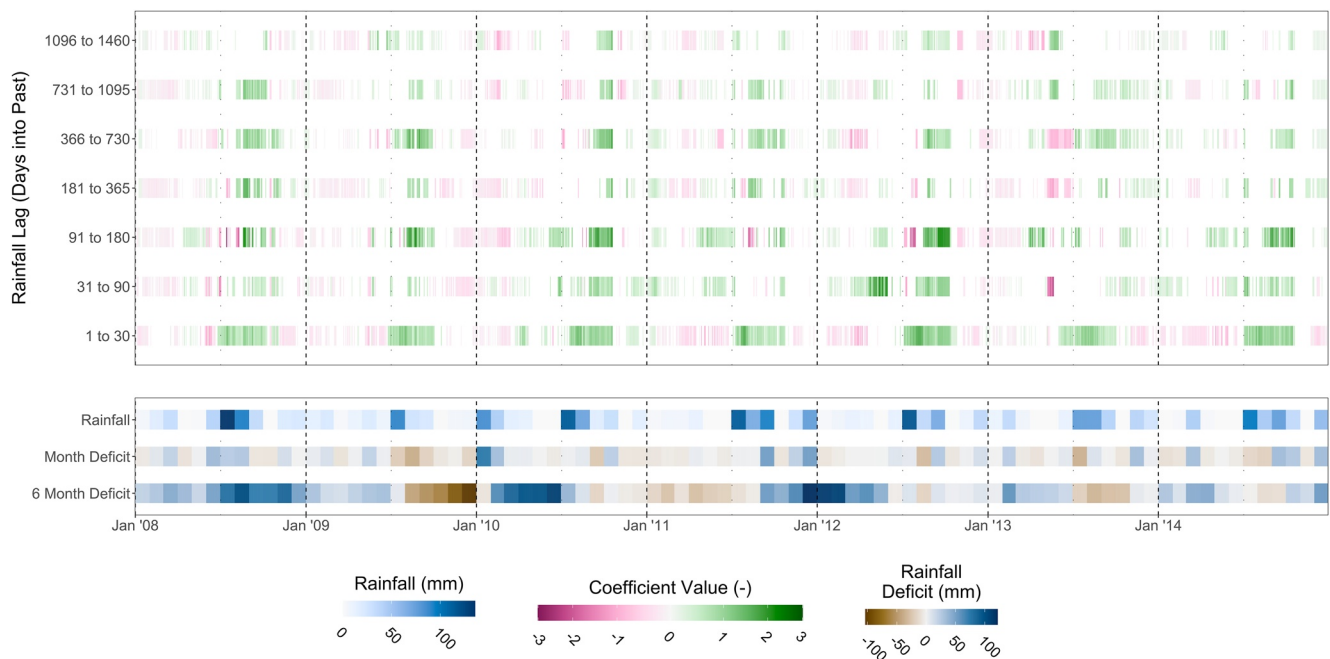
Figure 6. Timeseries of coefficient values for antecedent rainfall from the *k*-means clustering plus regression historical model for net ecosystem productivity at US-SRM. The plot layout and definitions are identical to Figure 5.

confounding influence of irrigation or crop type. Finally, Figure 8 shows the regression coefficients for a woody savanna, US-Ton. This site shows extremely strong sensitivity to antecedent rainfall during the dry summer growing seasons (see figure scale and note magnitudes relative to the previous three figures). This sensitivity is however noticeably less strong during the very wet years of 2006 and 2011.

A potential confounding factor in our methodology is the temporal autocorrelation of precipitation at sites. If different periods of lagged rainfall are highly correlated at a site, then the method could arbitrarily divide the
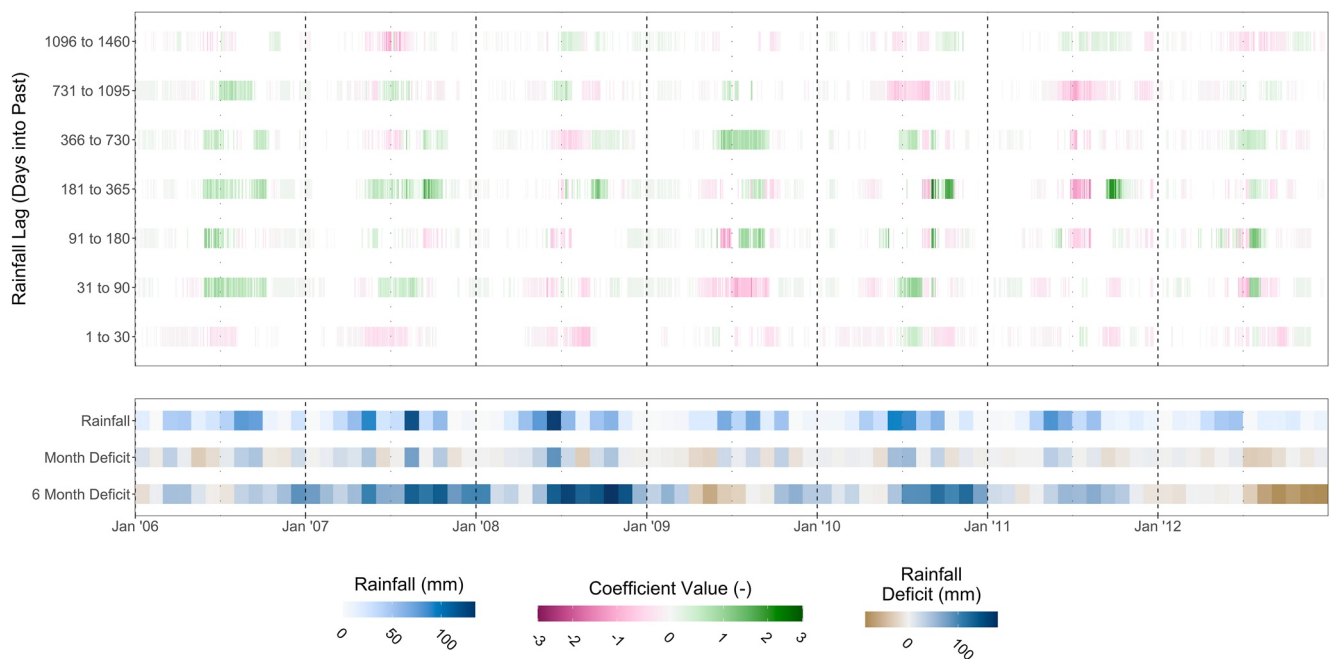

Figure 7. Timeseries of coefficient values for antecedent rainfall from the *k*-means clustering plus regression historical model for net ecosystem productivity at US-Ne3. The plot layout and definitions are identical to Figure 5.
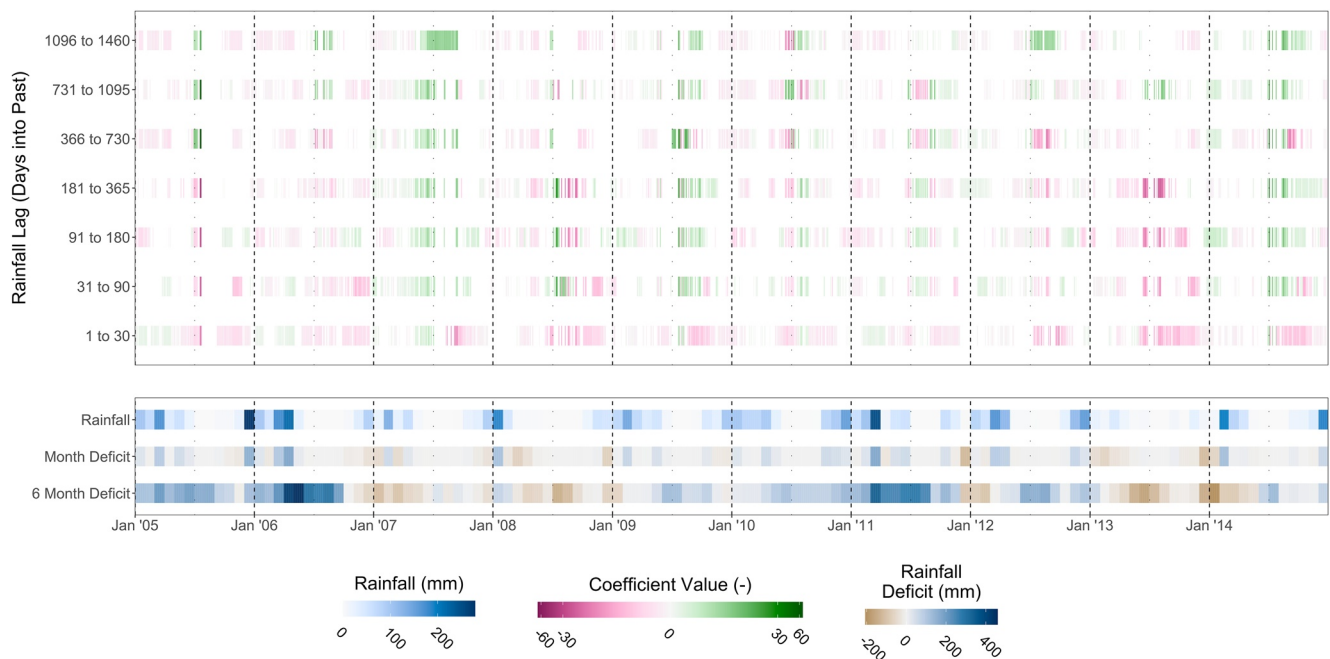
Figure 8. Timeseries of coefficient values for antecedent rainfall from the *k*-means clustering plus regression historical model for net ecosystem productivity at US-Ton. The plot layout and definitions are identical to Figure 5.

predictive power of the periods between them. This would result in a potential misclassification of sensitivity to one period of lagged climate over another. Of the 65 sites in this study, 37 of them exhibit at least one significant correlation ($p < 0.05$) between two periods of lagged rainfall. At BE-Lon, nine out of a possible 28 correlations between rainfall predictors are significantly correlated, while for the other 36 sites, at most 5 correlations are significant with 31 sites only having two or less significant correlations. 21 sites have significant correlations between lagged rainfall experienced between 31 and 90 days into the past, and rainfall from 181 to 365 days into the past. This is likely due to yearly rainfall seasonality, with one of these windows consisting of the site's dry/wet season and the other then being the inverse.

Nevertheless, at no site is current day precipitation significantly correlated with any lagged rainfall predictor, indicating that attribution of flux sensitivity to memory is unlikely to be affected. In effect therefore, any confounding effects are limited to the timescales at which this memory acts.

4. Discussion

Our machine learning methodology demonstrates the capacity to accurately model the variability in fluxes across a wide range of climates and vegetation types, as shown by the performance metrics in Table 2. Further, we have demonstrated the value of the methodology as a method for data mining flux data that does not require any a priori assumptions or knowledge regarding extreme climate events or specific lags. Similarly, this framework could prove an excellent tool for initial exploration of experimental data where ecosystem memory is hypothesized to be important. The results also imply an important role for plant functional type in determining the memory of an ecosystem to varying timescales of antecedent rainfall. For instance, forests had consistent NEP sensitivity to lagged rainfall received any time in the past 4 years, while savannah NEP was negatively sensitive to seasonal and mid-term antecedent rainfall but positively affected by long-term antecedent rainfall. Future work that attempts to mechanistically link vegetation traits to meteorological legacies is therefore likely valuable.

Theoretically our framework would highlight both lags and legacies in the response of fluxes to antecedent rainfall (and other drivers if included). In fact, the individual site plots in Figures 5–8 do show the ability of machine learning frameworks to identify how climate extremes affect flux sensitivity, with clear signals in the memory to precipitation during dry years. In addition, it is possible that the sensitivities at longer timescales (i.e., the "long-term memory") are driven by legacy mechanisms due to the length of time involved. However, it is notable that

well-known climate extremes (such as the 2003 European drought) were not explicitly identifiable in the time-series of lagged rainfall sensitivity at sites known to be affected. If, as expected, these well-known extreme events did significantly impact the site fluxes, it might be expected that the event would propagate through the timeseries of lagged rainfall. For instance, for the 2003 European drought which occurred in July-August, increased sensitivity to lagged rainfall might be expected in the lags that incorporate this time period (i.e., increased sensitivity in the "31–90 Day Lag" between August and October 2003, in the "91–180 Day Lag" between October 2003 and January 2004, and so on until the "1,096–1,460 Day Lag" which would respond between August 2006 and August 2007). There are numerous reasons why such behavior is not apparent (see Figures S1–S65 in Supporting Information S1 for European site timeseries). First, this could be a consequence of the framework used in this study. As the lagged predictors become more long-term, they contain longer periods of aggregated precipitation, and so the signal from a climate extreme may be lost (e.g., while July-August was exceptionally hot and dry in 2003, the 12 months both preceding and following did not experience particularly low rainfall). This could be tested in further work using more narrow lagged precipitation windows such that the climate extremes are more likely to be isolated and not be averaged out. Alternatively, it may well be the case that legacies to drought, while apparent in growth patterns (e.g., tree rings), are not evident in carbon and water fluxes (i.e., source limitations, Kannenberg et al., 2019). Our results could potentially provide additional evidence that the fluxes at drought-affected sites quickly recover to pre-drought levels, although additional work would be required as mentioned above to rule out other reasons for the lack of unambiguous legacies in our results. Applying our approach to both flux and tree ring data at overlapping sites would be a useful next step.

We now revisit the hypotheses outlined at the beginning of the paper, namely that (a) the inclusion of ecosystem memory improves the accuracy of modeled fluxes; (b) the sensitivity of fluxes to antecedent climate is non-stationary and varies both intra- and inter-annually; (c) ecosystem memory is more influential in the modeling of NEP than LE due to NEP having more mechanisms for delayed responses to prior conditions; and (d) the more arid a site is, the greater improvement in model performance when antecedent climate is included.

4.1. Does Memory Matter?

Understanding the role of lags and/or legacies is critical for accurate modeling and ensuring predictions of future climate change are properly constrained (Humphrey et al., 2018; Keenan et al., 2012; Ogle et al., 2015). Our results suggest that the importance of understanding vegetation response to prior climate is strongly site dependent. When all sites are viewed collectively, the introduction of memory effects to precipitation and temperature into our model improved performance metrics by between 7% and 24%. This supports the recent calls for a more nuanced inclusion of memory into LSMs (Anderegg et al., 2015b; Bastos et al., 2021; Frank et al., 2015; Jones et al., 2020; Kolus et al., 2019). However, the improvement varied substantially between sites and across time. BE-Lon, a cropland in Belgium, saw relative improvements in NEP modeling by 56% for r^2 (0.58 instantaneous to 0.91 historical) and 52% for NME (0.57 instantaneous to 0.27 historical). By contrast, an evergreen needleleaf forest in Canada, CA-Qfo, had no significant difference between the instantaneous and historical models for NEP (r^2 relative improvement of 0%, NME relative improvement of 4%). These improvements fall within the ranges reported in previous analysis (Cranko Page et al., 2022; Y. Liu et al., 2019).

Our results also show that lags and legacies have different impacts depending on the flux being modeled. Modeled NEP fluxes generally benefitted more from the introduction of antecedent climate than LE fluxes, consistent with a prior study (Cranko Page et al., 2022). This apparent discrepancy between the role of memory in the modeling of fluxes is due in part to the better performance of the instantaneous models for LE (more predictable) resulting in a lower potential benefit of memory effects. This implies that LE fluxes are driven mostly by the current climate (Best et al., 2015; Haughton, Abramowitz, & Pitman, 2018). Such a result is not surprising, since LE includes all components of the water flux (i.e., soil evaporation, canopy evaporation and transpiration). While partitioning of LE into evaporation and transpiration is difficult (Stoy et al., 2019), non-plant evaporation is estimated to account for around one-third of global ET (Fatichi & Pappas, 2017; Lian et al., 2018; Schlesinger & Jasechko, 2014; Wei et al., 2017). In energy-limited ecosystems (AI < 1), evaporation is driven predominately by concurrent radiation, temperature, wind speed and vapor pressure deficit (Monteith, 1965; Priestley & Taylor, 1972). In arid ecosystems, evaporation is more likely to depend on antecedent rainfall to ensure there is moisture available. While evaporation is clearly inherently instantaneous (when water and energy are available), transpiration, particularly from deeper-rooted species, could potentially provide mechanisms for memory to act

(Pérez-Ruiz et al., 2022). For instance, delays between rainfall and peak plant water content could buffer the response of transpiration to precipitation (Feldman et al., 2020). Indeed, it has been hypothesized that access to groundwater is a key driver of the interannual variability in plant functioning (Humphrey et al., 2018). Memory in the LE flux acting through transpiration of long-term water storage reservoirs is potentially the mechanism explaining why certain forest sites in our study (such as IT-SRo, FR-Pue, and AU-Ctr) see a relatively greater memory influence in LE compared to NEP fluxes compared to other sites. Therefore, there are clear mechanisms explaining the dominant role of current climate in LE fluxes, and hence the good performance of the instantaneous LE model. By illustrating differences in the memory of NEP and LE fluxes, we have contributed further evidence that memory effects are inconsistent across ecosystem fluxes and types (Kannenberg et al., 2019, 2020).

4.2. What Influences the Role of Memory at a Site?

Previous work has suggested that memory effects become more influential in increasingly arid ecosystems (Cranko Page et al., 2022; X. Liu et al., 2018. Y. Liu et al., 2019; D. Wu et al., 2015). Here, we find that the role of lags and legacies in ecosystem fluxes is not well explained by site aridity. There is no correlation between improvement in model performance for the historical model and either AI or EI at the site. This is unlikely to be due to our site selection, as we have 30 sites with AI greater than one (water-limited) and 35 with AI less than one (energy-limited). The sites studied have good coverage between AI values of 0.3 and 2 and include nine sites with AI over 2. However, some sites do lie far from the Budyko curve (Figure 2). This indicates that are potentially sources of water other than rainfall which may impact the ability of the models to accurately capture these sites (Haughton, Abramowitz, De Kauwe, & Pitman, 2018). For instance, our study includes cropland sites with no differentiation between irrigated and rainfed sites. This is a potential confounding factor when discussing these sites, as the total water available to the plants at an irrigated site is not represented accurately by the AI. Note however that memory importance is also not correlated with the distance of the site from the Budyko curve. One reason for such a difference in the role of aridity may be the ability of our model to allow sensitivity to precipitation memory to vary over time. Arid sites tend to experience less frequent rainfall and typically have shallow rooting systems (Schenk & Jackson, 2002) and the fluxes are therefore more tightly linked to antecedent rainfall, a proxy for soil moisture availability. The influence of precipitation memory at these sites would then be relatively constant. One surprising finding from our analysis is the generally low sensitivity of forest ecosystem fluxes to the longer timescales of rainfall. One interpretation of this result may be that deep soil water storage/ root access (including groundwater dependent ecosystems) blurs the exact timescales of influence of past rainfall. Thus, while overall there remains a sensitivity to past conditions in the flux predictability, it can be harder to attribute to a distinct timescale. Alternatively, this finding may imply that ecosystem fluxes themselves are relatively robust to past climatic conditions, as distinct from impacts on state conditions (*e.g.,* growth, leaf area; Sala et al., 2012). Understanding why memory effects vary so much at sites with different IGBP classifications (Figure 4) will likely require greater insight into species differences and site characteristics (*e.g.,* soil, groundwater access, species). For LE, sensitivity to antecedent rainfall at nearly all IGBP classifications was skewed positive (an increase in prior rainfall leads to an increase in LE) and decreased as the length of the lag increased. This provides further evidence of the role of evaporation, as increases in antecedent rainfall would increase soil moisture and surface runoff, leading to greater evaporation and LE. In contrast, NEP sensitivity to antecedent precipitation was more symmetrically distributed between positive and negative sensitivity. The greater magnitudes of sensitivity to antecedent precipitation seen at grassy sites would seem to support this positive/negative variability in sensitivity, suggesting a potential "boom-bust" behavior where high rainfall results in strong carbon uptake for grasses and, in dry years, carbon is released back to the atmosphere via respiration.

4.3. Does Memory Sensitivity Change Through Time?

Most studies of ecosystem memory utilize methods that assume, explicitly or implicitly, that the timescales of influence are constant at a site, both inter- and intra-annually. Such methods include regressions or correlation calculations against lagged rainfall (e.g., Richard et al., 2008; Hovenden et al., 2014; X. Liu et al., 2018) or frameworks such as Stochastic Antecedent Modeling (e.g., Barron-Gafford et al., 2014; Peltier & Ogle, 2019). However, it is reasonable to hypothesize that the role of memory might vary both throughout a year (for instance, during the growing season when plant growth provides more mechanisms by which NEP can be influenced by prior water) and across years (such as a particularly dry year dramatically reducing plant growth and therefore limiting any influence from prior rainfall).

Figures 5–8 demonstrate a key strength of our method and clearly show how the role of memory is non-stationary and changes through time. First, we see clear seasonal signals at some sites where the role of memory is greatest during (or even effectively limited to) a certain time of year, whether that be the growing season (US-Ne3 and US-SRM) or the dry season (US-Var and US-Ton). By assuming any memory effect is stationary, these seasonal signals may be masked by the lower—or even reversed—sensitivity at other times of year. Additionally, we show how different conditions can affect the role of memory. For instance, US-Var shows the strongest memory sensitivities in the driest years (2007, 2008, and 2013, see Figure 5). This strong negative sensitivity at a grassland implies that an increase in antecedent rainfall led to a decrease in the carbon sink at the site, which again is potentially due to a "boom/bust scenario." As the site becomes drier throughout the year, any carbon that was invested in additional biomass under antecedent conditions can no longer be supported, the plants turnover tissue, and the carbon is respired back into the atmosphere. Further evidence for this being a key mechanism in the negative sensitivity of NEP to antecedent rainfall is that the ecosystems where such behavior is known to occur most frequently (grasslands/savannas) are part of the PFTs which show the strongest negative sensitivities. There are other years (e.g., 2012) where there is almost no effect of antecedent rainfall on NEP. As such, we have shown that ecosystem memory varies in time and should not be assumed to be stationary. Any methods exploring the role of antecedent climate should consider the impact of changes in sensitivity through time at both inter- and intra-annual timescales. Our methodology is likely to be widely applicable when examining the impact of stochastic disturbances (e.g., fires, wind throw, pest attacks) and meteorological extremes (i.e., drought/heatwaves), as well as looking at the changing nature of vegetation response to climate as flux records increase in length. Identifying methods to summarize the time-varying memory effects across multiple sites is not trivial and is a direction for future work to ensure that the information from this framework is as accessible as possible.

This capacity to isolate the timescales of influence of both the past extremes and behavioral lags opens important avenues around the introduction of new theory into LSMs to capture ecosystem memory to climate. Our framework provides an approach for important checks on model hypothesis testing around the introduction of new theory related to plant hydraulics (De Kauwe et al., 2022; Sabot et al., 2020, 2022), acclimation (Mercado et al., 2018; Smith & Dukes, 2013), and carbon storage (De Kauwe et al., 2014; Fatichi et al., 2014; Jones et al., 2020). Implementing model hypotheses intended to improve the response timescales of vegetation to climate into LSMs, and then comparing the results between the model and observations when passed through frameworks such as ours, will ensure that model development is correctly capturing the timescales of influence that are evident in the observational records. Using this *k*-means clustering plus regression approach on testing sites used for model evaluation could help to benchmark expected flux responses to antecedent climate. This link between machine learning and model theory development is an important next step for LSMs.

5. Conclusion

We showed that memory effects can play a significant role in the modeling of ecosystem fluxes and functioning. However, our results also indicate that this effect is highly variable and that some sites do not require a memory component to explain the measured fluxes at all. In contrast to previous work, we find that aridity is not a good predictor for whether a site exhibits strong memory effects. Further, our method provided evidence that the role of memory can change through time and should not be assumed to be stationary. It provides a simple tool to site investigators to better explain the drivers influencing ecosystem functioning in different circumstances. Such exploration could unlock further understanding of plant functioning and ecosystem mechanisms that govern the carbon and water cycle. This understanding is in turn fundamental to improving the LSMs that are used to predict the climate of the future and the impacts of anthropogenic climate change.

Data Availability Statement

The FLUXNET data are available at https://doi.org/10.25914/5fdb0902607e1 (Ukkola, 2021). Examples of model and analysis codes are available at https://doi.org/10.5281/zenodo.7527607 (Cranko Page, 2023).

References

Abramowitz, G., Leuning, R., Clark, M., & Pitman, A. (2008). Evaluating the performance of land surface models. *Journal of Climate*, *21*, 5468–5481. https://doi.org/10.1175/2008JCLI2378.1

Alexander, L. V. (2016). Global observed long-term changes in temperature and precipitation extremes: A review of progress and limitations in IPCC assessments and beyond. *Weather and Climate Extremes*, *11*, 4–16. https://doi.org/10.1016/j.wace.2015.10.007

Allan, R. P., & Soden, B. J. (2008). Atmospheric warming and the amplification of precipitation extremes. *Science*, *321*(5895), 1481–1484. https://doi.org/10.1126/science.1160787

Allen, C. D., Breshears, D. D., & McDowell, N. G. (2015). On underestimation of global vulnerability to tree mortality and forest die-off from hotter drought in the Anthropocene. *Ecosphere*, *6*(8), 129. https://doi.org/10.1890/ES15-00203.1

Alves, R. D. F. B., Menezes-Silva, P. E., Sousa, L. F., Loram-Lourenço, L., Silva, M. L. F., Almeida, S. E. S., et al. (2020). Evidence of drought memory in dipteryx alata indicates differential acclimation of plants to savanna conditions. *Scientific Reports*, *10*(1), 16455. https://doi.org/10.1038/s41598-020-73423-3

Anderegg, W. R. L., Flint, A., Huang, C., Flint, L., Berry, J. A., Davis, F. W., et al. (2015a). Tree mortality predicted from drought-induced vascular damage. *Nature Geoscience*, *8*(5), 5. https://doi.org/10.1038/ngeo2400

Anderegg, W. R. L., Schwalm, C., Biondi, F., Camarero, J. J., Koch, G., Litvak, M., et al. (2015b). Pervasive drought legacies in forest ecosystems and their implications for carbon cycle models. *Science*, *349*(6247), 528–532. https://doi.org/10.1126/science.aab1833

Arca, V., Power, S. A., Delgado-Baquerizo, M., Pendall, E., & Ochoa-Hueso, R. (2021). Seasonal effects of altered precipitation regimes on ecosystem-level CO2 fluxes and their drivers in a grassland from Eastern Australia. *Plant and Soil*, *460*(1–2), 435–451. https://doi.org/10.1007/s11104-020-04811-x

Arend, M., Link, R. M., Patthey, R., Hoch, G., Schuldt, B., & Kahmen, A. (2021). Rapid hydraulic collapse as cause of drought-induced mortality in conifers. *Proceedings of the National Academy of Sciences*, *118*(16), e2025251118. https://doi.org/10.1073/pnas.2025251118

Arthur, D., & Vassilvitskii, S. (2007). K-Means++: The advantages of careful seeding. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms* (pp. 1027–1035). https://doi.org/10.1145/1283383.1283494

Barron-Gafford, G. A., Cable, J. M., Bentley, L. P., Scott, R. L., Huxman, T. E., Jenerette, G. D., & Ogle, K. (2014). Quantifying the timescales over which exogenous and endogenous conditions affect soil respiration. *New Phytologist*, *202*(2), 442–454. https://doi.org/10.1111/nph.12675

Bastos, A., Ciais, P., Friedlingstein, P., Sitch, S., Pongratz, J., Fan, L., et al. (2020). Direct and seasonal legacy effects of the 2018 heat wave and drought on European ecosystem productivity. *Science Advances*, *6*(24), eaba2724. https://doi.org/10.1126/sciadv.aba2724

Bastos, A., Orth, R., Reichstein, M., Ciais, P., Viovy, N., Zaehle, S., et al. (2021). Vulnerability of European ecosystems to two compound dry and hot summers in 2018 and 2019. *Earth System Dynamics*, *12*(4), 1015–1035. https://doi.org/10.5194/esd-12-1015-2021

Beguería, S., Vicente-Serrano, S. M., Reig, F., & Latorre, B. (2014). Standardized precipitation evapotranspiration index (SPEI) revisited: Parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *International Journal of Climatology*, *34*(10), 3001–3023. https://doi.org/10.1002/joc.3887

Best, M. J., Abramowitz, G., Johnson, H. R., Pitman, A. J., Balsamo, G., Boone, A., et al. (2015). The plumbing of land surface models: Benchmarking model performance. *Journal of Hydrometeorology*, *16*(3), 1425–1442. https://doi.org/10.1175/JHM-D-14-0158.1

Bigler, C., Gavin, D. G., Gunning, C., & Veblen, T. T. (2007). Drought induces lagged tree mortality in a subalpine forest in the Rocky Mountains. *Oikos*, *116*(12), 1983–1994. https://doi.org/10.1111/j.2007.0030-1299.16034.x

Budyko, M. I. (1974). *Climate and life, International geophysics series*. Academic Press.

Cable, J. M., Ogle, K., Barron-Gafford, G. A., Bentley, L. P., Cable, W. L., Scott, R. L., et al. (2013). Antecedent conditions influence soil respiration differences in shrub and grass patches. *Ecosystems*, *16*(7), 1230–1247. https://doi.org/10.1007/s10021-013-9679-7

Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogée, J., Allard, V., et al. (2005). Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature*, *437*(7058), 529–533. https://doi.org/10.1038/nature03972

Cleverly, J., Boulain, N., Villalobos-Vega, R., Grant, N., Faux, R., Wood, C., et al. (2013). Dynamics of component carbon fluxes in a semi-arid Acacia woodland, central Australia. *Journal of Geophysical Research: Biogeosciences*, *118*(3), 1168–1185. https://doi.org/10.1002/jgrg.20101

Cleverly, J., Eamus, D., Restrepo Coupe, N., Chen, C., Maes, W., Li, L., et al. (2016). Soil moisture controls on phenology and productivity in a semi-arid critical zone. *Science of the Total Environment*, *568*, 1227–1237. https://doi.org/10.1016/j.scitotenv.2016.05.142

Cranko Page, J. (2023). kmeans_memory. https://doi.org/10.5281/zenodo.7527607

Cranko Page, J., De Kauwe, M. G., Abramowitz, G., Cleverly, J., Hinko-Najera, N., Hovenden, M. J., et al. (2022). Examining the role of environmental memory in the predictability of carbon and water fluxes across Australian ecosystems. *Biogeosciences*, *19*(7), 1913–1932. https://doi.org/10.5194/bg-19-1913-2022

De Boeck, H. J., Hiltbrunner, E., Verlinden, M., Bassin, S., & Zeiter, M. (2018). Legacy effects of climate extremes in alpine grassland. *Frontiers of Plant Science*, *9*. https://doi.org/10.3389/fpls.2018.01586

De Kauwe, M. G., Medlyn, B. E., Zaehle, S., Walker, A. P., Dietze, M. C., Wang, Y.-P., et al. (2014). Where does the carbon go? A model–data intercomparison of vegetation carbon allocation and turnover processes at two temperate forest free-air CO$_2$ enrichment sites. *New Phytologist*, *203*(3), 883–899. https://doi.org/10.1111/nph.12847

De Kauwe, M. G., Sabot, M. E. B., Medlyn, B. E., Pitman, A. J., Meir, P., Cernusak, L. A., et al. (2022). Towards species-level forecasts of drought-induced tree mortality risk. *New Phytologist*, *235*(1), 94–110. https://doi.org/10.1111/nph.18129

Dunn, R. J. H., Alexander, L. V., Donat, M. G., Zhang, X., Bador, M., Herold, N., et al. (2020). Development of an updated global land in situ-based data set of temperature and precipitation extremes: HadEX3. *Journal of Geophysical Research: Atmospheres*, *125*(16), e2019JD032263. https://doi.org/10.1029/2019JD032263

Fatichi, S., Leuzinger, S., & Körner, C. (2014). Moving beyond photosynthesis: From carbon source to sink-driven vegetation modeling. *New Phytologist*, *201*(4), 1086–1095. https://doi.org/10.1111/nph.12614

Fatichi, S., & Pappas, C. (2017). Constrained variability of modeled T:ET ratio across biomes. *Geophysical Research Letters*, *44*(13), 6795–6803. https://doi.org/10.1002/2017GL074041

Feldman, A. F., Short Gianotti, D. J., Konings, A. G., Gentine, P., & Entekhabi, D. (2020). Patterns of plant rehydration and growth following pulses of soil moisture availability. *Biogeosciences Discussions*, 1–24. https://doi.org/10.5194/bg-2020-380

FLUXNET. (2023). *List of FLUXNET 2015 sites*. FLUXNET. Retrieved from https://fluxnet.org/sites/site-list-and-pages/

Frank, D., Reichstein, M., Bahn, M., Thonicke, K., Frank, D., Mahecha, M. D., et al. (2015). Effects of climate extremes on the terrestrial carbon cycle: Concepts, processes and potential future impacts. *Global Change Biology*, *21*(8), 2861–2880. https://doi.org/10.1111/gcb.12916

Gong, Y.-H., Zhao, D.-M., Ke, W.-B., Fang, C., Pei, J.-Y., Sun, G.-J., & Ye, J.-S. (2020). Legacy effects of precipitation amount and frequency on the aboveground plant biomass of a semi-arid grassland. *Science of the Total Environment*, *705*, 135899. https://doi.org/10.1016/j.scitotenv.2019.135899

Griffin-Nolan, R. J., Carroll, C. J. W., Denton, E. M., Johnston, M. K., Collins, S. L., Smith, M. D., & Knapp, A. K. (2018). Legacy effects of a regional drought on aboveground net primary production in six central US grasslands. *Plant Ecology*, *219*(5), 505–515. https://doi.org/10.1007/s11258-018-0813-7

Guo, J. S., Gear, L., Hultine, K. R., Koch, G. W., & Ogle, K. (2020). Non-structural carbohydrate dynamics associated with antecedent stem water potential and air temperature in a dominant desert shrub. *Plant, Cell and Environment*, *43*(6), 1467–1483. https://doi.org/10.1111/pce.13749

Hahn, C., Lüscher, A., Ernst-Hasler, S., Suter, M., & Kahmen, A. (2021). Timing of drought in the growing season and strong legacy effects determine the annual productivity of temperate grasses in a changing climate. *Biogeosciences*, *18*(2), 585–604. https://doi.org/10.5194/bg-18-585-2021

Haughton, N., Abramowitz, G., De Kauwe, M. G., & Pitman, A. J. (2018). Does predictability of fluxes vary between FLUXNET sites? *Biogeosciences*, *15*(14), 4495–4513. https://doi.org/10.5194/bg-15-4495-2018

Haughton, N., Abramowitz, G., & Pitman, A. J. (2018). On the predictability of land surface fluxes from meteorological variables. *Geoscientific Model Development*, *11*(1), 195–212. https://doi.org/10.5194/gmd-11-195-2018

Hovenden, M. J., Leuzinger, S., Newton, P. C. D., Fletcher, A., Fatichi, S., Lüscher, A., et al. (2019). Globally consistent influences of seasonal precipitation limit grassland biomass response to elevated CO_2. *Nature Plants*, *5*, 167–173. https://doi.org/10.1038/s41477-018-0356-x

Hovenden, M. J., Newton, P. C. D., & Newton, P. C. D. (2018). Variability in precipitation seasonality limits grassland biomass responses to rising CO_2: Historical and projected climate analyses. *Climatic Change*, *149*, 219–231. https://doi.org/10.1007/s10584-018-2227-x

Hovenden, M. J., Newton, P. C. D., & Wills, K. E. (2014). Seasonal not annual rainfall determines grassland biomass response to carbon dioxide. *Nature*, *511*(7511), 583–586. https://doi.org/10.1038/nature13281

Huang, M., Wang, X., Keenan, T. F., & Piao, S. (2018). Drought timing influences the legacy of tree growth recovery. *Global Change Biology*, *24*(8), 3546–3559. https://doi.org/10.1111/gcb.14294

Huang, M., Zhai, P., & Piao, S. (2021). Divergent responses of ecosystem water use efficiency to drought timing over Northern Eurasia. *Environmental Research Letters*, *14*(4), 045016. https://doi.org/10.1088/1748-9326/abf0d1

Humphrey, V., Zscheischler, J., Ciais, P., Gudmundsson, L., Sitch, S., & Seneviratne, S. I. (2018). Sensitivity of atmospheric CO_2 growth rate to observed changes in terrestrial water storage. *Nature*, *560*(7720), 628–631. https://doi.org/10.1038/s41586-018-0424-4

Huxman, T. E., Snyder, K. A., Tissue, D., Leffler, A. J., Ogle, K., Pockman, W. T., et al. (2004). Precipitation pulses and carbon fluxes in semiarid and arid ecosystems. *Oecologia*, *141*(2), 254–268. https://doi.org/10.1007/s00442-004-1682-4

IPCC. (2021). In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, et al. (Eds.), *Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge University Press. https://doi.org/10.1017/9781009157896

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, *31*(3), 264–323. https://doi.org/10.1145/331499.331504

Jiménez-Muñoz, J. C., Mattar, C., Barichivich, J., Santamaría-Artigas, A., Takahashi, K., Malhi, Y., et al. (2016). Record-breaking warming and extreme drought in the Amazon rainforest during the course of El Niño 2015–2016. *Scientific Reports*, *6*(1), 33130. https://doi.org/10.1038/srep33130

Jones, S., Rowland, L., Cox, P., Hemming, D., Wiltshire, A., Williams, K., et al. (2020). The impact of a simple representation of non-structural carbohydrates on the simulated response of tropical forests to drought. *Biogeosciences*, *17*(13), 3589–3612. https://doi.org/10.5194/bg-17-3589-2020

Kannenberg, S. A., Cabon, A., Babst, F., Belmecheri, S., Delpierre, N., Guerrieri, R., et al. (2022). Drought-induced decoupling between carbon uptake and tree growth impacts forest carbon turnover time. *Agricultural and Forest Meteorology*, *322*, 108996. https://doi.org/10.1016/j.agrformet.2022.108996

Kannenberg, S. A., Novick, K. A., Alexander, M. R., Maxwell, J. T., Moore, D. J. P., Phillips, R. P., & Anderegg, W. R. L. (2019). Linking drought legacy effects across scales: From leaves to tree rings to ecosystems. *Global Change Biology*, *25*(9), 2978–2992. https://doi.org/10.1111/gcb.14710

Kannenberg, S. A., Schwalm, C. R., & Anderegg, W. R. L. (2020). Ghosts of the past: How drought legacy effects shape forest functioning and carbon cycling. *Ecol Lett ele*, *13485*(5), 891–901. https://doi.org/10.1111/ele.13485

Keenan, T. F., Baker, I., Barr, A., Ciais, P., Davis, K., Dietze, M., et al. (2012). Terrestrial biosphere model performance for inter-annual variability of land-atmosphere CO_2 exchange. *Global Change Biology*, *18*(6), 1971–1987. https://doi.org/10.1111/j.1365-2486.2012.02678.x

Kolus, H. R., Huntzinger, D. N., Schwalm, C. R., Fisher, J. B., McKay, N., Fang, Y., et al. (2019). Land carbon models underestimate the severity and duration of drought's impact on plant productivity. *Scientific Reports*, *9*(1), 2758. https://doi.org/10.1038/s41598-019-39373-1

Lemoine, N. P., Griffin-Nolan, R. J., Lock, A. D., & Knapp, A. K. (2018). Drought timing, not previous drought exposure, determines sensitivity of two shortgrass species to water stress. *Oecologia*, *188*(4), 965–975. https://doi.org/10.1007/s00442-018-4265-5

Lian, X., Piao, S., Huntingford, C., Li, Y., Zeng, Z., Wang, X., et al. (2018). Partitioning global land evapotranspiration using CMIP5 models constrained by observations. *Nature Climate Change*, *8*(7), 640–646. https://doi.org/10.1038/s41558-018-0207-9

Liu, L., Zhang, Y., Wu, S., Li, S., & Qin, D. (2018). Water memory effects and their impacts on global vegetation productivity and resilience. *Scientific Reports*, *8*(1), 2962. https://doi.org/10.1038/s41598-018-21339-4

Liu, Y., Schwalm, C. R., Samuels-Crow, K. E., & Ogle, K. (2019). Ecological memory of daily carbon exchange across the globe and its importance in drylands. *Ecology Letters*, *22*(11), 1806–1816. https://doi.org/10.1111/ele.13363

Ma, X., Huete, A., Cleverly, J., Eamus, D., Chevallier, F., Joiner, J., et al. (2016). Drought rapidly diminishes the large net CO 2 uptake in 2011 over semi-arid Australia. *Scientific Reports*, *6*(1), 37747. https://doi.org/10.1038/srep37747

MacQueen, J. (1967). Classification and analysis of multivariate observations. In *Fifth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 281–297).

Mantoan, L. P. B., Corrêa, C. V., Rainho, C. A., & de Almeida, L. F. R. (2020). Rapid dehydration induces long-term water deficit memory in sorghum seedlings: Advantages and consequences. *Environmental and Experimental Botany*, *180*, 104252. https://doi.org/10.1016/j.envexpbot.2020.104252

Mercado, L. M., Medlyn, B. E., Huntingford, C., Oliver, R. J., Clark, D. B., Sitch, S., et al. (2018). Large sensitivity in land carbon storage due to geographical and temporal variation in the thermal response of photosynthetic capacity. *New Phytologist*, *218*(4), 1462–1477. https://doi.org/10.1111/nph.15100

Min, S.-K., Zhang, X., Zwiers, F. W., & Hegerl, G. C. (2011). Human contribution to more-intense precipitation extremes. *Nature*, *470*(7334), 7334–7381. https://doi.org/10.1038/nature09763

Monteith, J. L. (1965). Evaporation and environment. In *Symposia of the Society for Experimental Biology* (Vol. 19, pp. 205–234).

Moran, M. S., Ponce-Campos, G. E., Huete, A., McClaran, M. P., Zhang, Y., Hamerlynck, E. P., et al. (2014). Functional response of U.S. grasslands to the early 21st-century drought. *Ecology*, *95*(8), 2121–2133. https://doi.org/10.1890/13-1687.1

Mouselimis, L. (2022). ClusterR: Gaussian mixture models, K-means, mini-batch-kmeans, K-medoids and affinity propagation clustering (manual).

Nearing, G. S., Ruddell, B. L., Clark, M. P., Nijssen, B., & Peters-Lidard, C. (2018). Benchmarking and process diagnostics of land models. *Journal of Hydrometeorology*, *19*(11), 1835–1852. https://doi.org/10.1175/JHM-D-17-0209.1

Ogle, K., Barber, J. J., Barron-Gafford, G. A., Bentley, L. P., Young, J. M., Huxman, T. E., et al. (2015). Quantifying ecological memory in plant and ecosystem processes. *Ecology Letters*, *18*(3), 221–235. https://doi.org/10.1111/ele.12399

Peltier, D. M. P., Barber, J. J., & Ogle, K. (2018). Quantifying antecedent climatic drivers of tree growth in the Southwestern US. *Journal of Ecology*, *106*(2), 613–624. https://doi.org/10.1111/1365-2745.12878

Peltier, D. M. P., Guo, J., Nguyen, P., Bangs, M., Wilson, M., Samuels-Crow, K., et al. (2021). Temperature memory and non-structural carbohydrates mediate legacies of a hot drought in trees across the southwestern US. *Tree Physiology*, *42*(1), 71–85. https://doi.org/10.1093/treephys/tpab091

Peltier, D. M. P., & Ogle, K. (2019). Legacies of more frequent drought in ponderosa pine across the western United States. *Global Change Biology*, *25*(11), 3803–3816. https://doi.org/10.1111/gcb.14720

Peltier, D. M. P., & Ogle, K. (2020). Tree growth sensitivity to climate is temporally variable. *Ecology Letters*, *23*(11), 1561–1572. https://doi.org/10.1111/ele.13575

Pérez-Ruiz, E. R., Vivoni, E. R., & Sala, O. E. (2022). Seasonal carryover of water and effects on carbon dynamics in a dryland ecosystem. *Ecosphere*, *13*(7), e4189. https://doi.org/10.1002/ecs2.4189

Perkins-Kirkpatrick, S. E., & Lewis, S. C. (2020). Increasing trends in regional heatwaves. *Nature Communications*, *11*(1), 3357. https://doi.org/10.1038/s41467-020-16970-7

Priestley, C. H. B., & Taylor, R. J. (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review*, *100*(2), 81–92. https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M. D., Seneviratne, S. I., et al. (2013). Climate extremes and the carbon cycle. *Nature*, *500*(7462), 287–295. https://doi.org/10.1038/nature12350

Renchon, A. A., Drake, J. E., Macdonald, C. A., Sihi, D., Hinko-Najera, N., Tjoelker, M. G., et al. (2021). Concurrent measurements of soil and ecosystem respiration in a mature eucalypt woodland: Advantages, lessons, and questions. *Journal of Geophysical Research: Biogeosciences*, *126*(3), e2020JG006221. https://doi.org/10.1029/2020JG006221

Richard, Y., Martiny, N., Fauchereau, N., Reason, C., Rouault, M., Vigaud, N., & Tracol, Y. (2008). Interannual memory effects for spring NDVI in semi-arid South Africa. *Geophysical Research Letters*, *35*(13), L13704. https://doi.org/10.1029/2008GL034119

Ryan, E. M., Ogle, K., Peltier, D., Walker, A. P., De Kauwe, M. G., Medlyn, B. E., et al. (2017). Gross primary production responses to warming, elevated CO_2, and irrigation: Quantifying the drivers of ecosystem physiology in a semiarid grassland. *Global Change Biology*, *23*(8), 3092–3106. https://doi.org/10.1111/gcb.13602

Ryan, E. M., Ogle, K., Zelikova, T. J., LeCain, D. R., Williams, D. G., Morgan, J. A., & Pendall, E. (2015). Antecedent moisture and temperature conditions modulate the response of ecosystem respiration to elevated CO_2 and warming. *Global Change Biology*, *21*(7), 2588–2602. https://doi.org/10.1111/gcb.12910

Sabot, M. E., De Kauwe, M. G., Pitman, A. J., Medlyn, B. E., Ellsworth, D. S., Martin-StPaul, N. K., et al. (2022). One stomatal model to rule them all? Toward improved representation of carbon and water exchange in global models. *Journal of Advances in Modeling Earth Systems*, *14*(4), e2021MS002761. https://doi.org/10.1029/2021ms002761

Sabot, M. E. B., Kauwe, M. G. D., Pitman, A. J., Medlyn, B. E., Verhoef, A., Ukkola, A. M., & Abramowitz, G. (2020). Plant profit maximization improves predictions of European forest responses to drought. *New Phytologist*, *226*(6), 1638–1655. https://doi.org/10.1111/nph.16376

Sala, O. E., Gherardi, L. A., Reichmann, L., Jobbágy, E., & Peters, D. (2012). Legacies of precipitation fluctuations on primary production: Theory and data synthesis. *Philosophical Transactions of the Royal Society B*, *367*(1606), 3135–3144. https://doi.org/10.1098/rstb.2011.0347

Schär, C., Vidale, P. L., Lüthi, D., Frei, C., Häberli, C., Liniger, M. A., & Appenzeller, C. (2004). The role of increasing temperature variability in European summer heatwaves. *Nature*, *427*(6972), 332–336. https://doi.org/10.1038/nature02300

Schenk, H. J., & Jackson, R. B. (2002). Rooting depths, lateral root spreads and below-ground/above-ground allometries of plants in water-limited ecosystems. *Journal of Ecology*, *90*(3), 480–494. https://doi.org/10.1046/j.1365-2745.2002.00682.x

Schlesinger, W. H., & Jasechko, S. (2014). Transpiration in the global water cycle. *Agricultural and Forest Meteorology*, *189–190*, 115–117. https://doi.org/10.1016/j.agrformet.2014.01.011

Schwalm, C. R., Anderegg, W. R. L., Michalak, A. M., Fisher, J. B., Biondi, F., Koch, G., et al. (2017). Global patterns of drought recovery. *Nature*, *548*(7666), 202–205. https://doi.org/10.1038/nature23021

Seabloom, E. W., Borer, E. T., & Tilman, D. (2020). Grassland ecosystem recovery after soil disturbance depends on nutrient supply rate. *European Journal of Neuroscience*, *13591*(12), 1756–1765. https://doi.org/10.1111/ele.13591

Smith, N. G., & Dukes, J. S. (2013). Plant respiration and photosynthesis in global-scale models: Incorporating acclimation to temperature and CO2. *Global Change Biology*, *19*(1), 45–63. https://doi.org/10.1111/j.1365-2486.2012.02797.x

Stott, P. A., Stone, D. A., & Allen, M. R. (2004). Human contribution to the European heatwave of 2003. *Nature*, *432*(7017), 610–614. https://doi.org/10.1038/nature03089

Stoy, P. C., El-Madany, T. S., Fisher, J. B., Gentine, P., Gerken, T., Good, S. P., et al. (2019). Reviews and syntheses: Turning the challenges of partitioning ecosystem evaporation and transpiration into opportunities. *Biogeosciences*, *16*(19), 3747–3775. https://doi.org/10.5194/bg-16-3747-2019

Strahler, A., Muchoney, D., Borak, J., Friedl, M., Gopal, S., Lambin, E., & Moody, A. (1999). MODIS land cover product algorithm theoretical basis document (ATBD) version 5.0. Retrieved from http://modis.gsfc.nasa.gov/data/atbd/atbd_mod12.pdf

Sun, Q., Meyer, W. S., Koerber, G. R., & Marschner, P. (2020). Rapid recovery of net ecosystem production in a semi-arid woodland after a wildfire. *Agricultural and Forest Meteorology*, *291*, 108099. https://doi.org/10.1016/j.agrformet.2020.108099

Szejner, P., Belmecheri, S., Ehleringer, J. R., & Monson, R. K. (2020). Recent increases in drought frequency cause observed multi-year drought legacies in the tree rings of semi-arid forests. *Oecologia*, *192*(1), 241–259. https://doi.org/10.1007/s00442-019-04550-6

Thornthwaite, C. W. (1948). An approach toward a rational classification of climate. *Geographical Review*, *38*(1), 55–94. https://doi.org/10.2307/210739

Ukkola, A. (2021). PLUMBER2 forcing and evaluation datasets for a model intercomparison project for land surface models. https://doi.org/10.25914/5fdb0902607e1

Ukkola, A. M., Abramowitz, G., & De Kauwe, M. G. (2022). A flux tower dataset tailored for land model evaluation. *Earth System Science Data*, *14*(2), 449–461. https://doi.org/10.5194/essd-14-449-2022

Ukkola, A. M., De Kauwe, M. G., Roderick, M. L., Burrell, A., Lehmann, P., & Pitman, A. J. (2021). Annual precipitation explains variability in dryland vegetation greenness globally but not locally. *Global Change Biology*, *27*(18), 4367–4380. https://doi.org/10.1111/gcb.15729

Vanoni, M., Bugmann, H., Nötzli, M., & Bigler, C. (2016). Quantifying the effects of drought on abrupt growth decreases of major tree species in Switzerland. *Ecology and Evolution*, *6*(11), 3555–3570. https://doi.org/10.1002/ece3.2146

Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *Journal of Climate*, *23*(7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1

Wei, Z., Yoshimura, K., Wang, L., Miralles, D. G., Jasechko, S., & Lee, X. (2017). Revisiting the contribution of transpiration to global terrestrial evapotranspiration. *Geophysical Research Letters*, *44*(6), 2792–2801. https://doi.org/10.1002/2016GL072235

Williams, A. P., Cook, B. I., & Smerdon, J. E. (2022). Rapid intensification of the emerging southwestern North American megadrought in 2020–2021. *Nature Climate Change*, *12*(3), 232–234. https://doi.org/10.1038/s41558-022-01290-z

Wu, D., Zhao, X., Liang, S., Zhou, T., Huang, K., Tang, B., & Zhao, W. (2015). Time-lag effects of global vegetation responses to climate change. *Global Change Biology*, *21*(9), 3520–3531. https://doi.org/10.1111/gcb.12945

Wu, X., Liu, H., Li, X., Ciais, P., Babst, F., Guo, W., et al. (2018). Differentiating drought legacy effects on vegetation growth over the temperate Northern Hemisphere. *Global Change Biology*, *24*(1), 504–516. https://doi.org/10.1111/gcb.13920

Zhang, T., Xu, M., Xi, Y., Zhu, J., Tian, L., Zhang, X., et al. (2015). Lagged climatic effects on carbon fluxes over three grassland ecosystems in China. *Journal of Plant Ecology*, *8*(3), 291–302. https://doi.org/10.1093/jpe/rtu026

Zhang, X., Wan, H., Zwiers, F. W., Hegerl, G. C., & Min, S.-K. (2013). Attributing intensification of precipitation extremes to human influence. *Geophysical Research Letters*, *40*(19), 5252–5257. https://doi.org/10.1002/grl.51010

Zscheischler, J., Mahecha, M. D., von Buttlar, J., Harmeling, S., Jung, M., Rammig, A., et al. (2014a). A few extreme events dominate global inter-annual variability in gross primary production. *Environmental Research Letters*, *9*(3), 035001. https://doi.org/10.1088/1748-9326/9/3/035001

Zscheischler, J., Michalak, A. M., Schwalm, C., Mahecha, M. D., Huntzinger, D. N., Reichstein, M., et al. (2014b). Impact of large-scale climate extremes on biospheric carbon fluxes: An intercomparison based on MsTMIP data. *Global Biogeochemical Cycles*, *28*(6), 585–600. https://doi.org/10.1002/2014GB004826