














Widespread underestimation of rain-induced soil carbon emissions from global drylands

Received: 24 October 2024

Accepted: 25 June 2025

Published online: 31 July 2025

 Check for updates

Ngoc B. Nguyen ¹✉, Mirco Migliavacca ², Maoya Bassiouni ¹,
Dennis D. Baldocchi ¹, Laureano A. Gherardi ¹, Julia K. Green ^{1,3},
Dario Papale ^{4,5}, Markus Reichstein ⁶, Kai-Hendrik Cohrs ⁷,
Alessandro Cescatti ², Tuan Dung Nguyen ⁸, Hoang H. Nguyen⁹,
Quang Minh Nguyen ¹⁰ & Trevor F. Keenan ^{1,11}✉

Dryland carbon fluxes, particularly those driven by ecosystem respiration, are highly sensitive to water availability and rain pulses. However, the magnitude of rain-induced carbon emissions remains unclear globally. Here we quantify the impact of rain-pulse events on the carbon balance of global drylands and characterize their spatiotemporal controls. Using eddy-covariance observations of carbon, water and energy fluxes from 34 dryland sites worldwide, we produce an inventory of over 1,800 manually identified rain-induced CO₂ pulse events. Based on this inventory, a machine learning algorithm is developed to automatically detect rain-induced CO₂ pulse events. Our findings show that existing partitioning methods underestimate ecosystem respiration and photosynthesis by up to 30% during rain-pulse events, which annually contribute $16.9 \pm 2.8\%$ of ecosystem respiration and $9.6 \pm 2.2\%$ of net ecosystem productivity. We show that the carbon loss intensity correlates most strongly with annual productivity, aridity and soil pH. Finally, we identify a universal decay rate of rain-induced CO₂ pulses and use it to bias-correct respiration estimates. Our research highlights the importance of rain-induced carbon emissions for the carbon balance of global drylands and suggests that ecosystem models may largely underrepresent the influence of rain pulses on the carbon cycle of drylands.

Drylands cover over a third of the global land surface^{1,2} and substantially influence the trend and interannual variability of the terrestrial carbon sink^{3–5}. These ecosystems are water limited, with rainfall driving vegetation and microbial processes that impact ecosystem carbon dynamics⁶. Sporadic rain pulses in particular play a crucial role in determining plant growth, microbial activity, soil moisture and overall ecosystem productivity^{7–9}. Although the influence of water availability on dryland carbon dynamics is well established^{6,10,11}, the processes governing responses to rain pulses remained poorly understood. The amount of carbon lost due to heterotrophic respiration stimulated by rain-pulse events and the key drivers of those losses across diverse dryland ecosystems, remain open questions^{12–15}, and the underlying processes are thus typically not included in models of dryland carbon dynamics^{14,16,17}.

Rain pulses trigger abiotic and biotic soil CO₂ pulses, a phenomenon often termed a rain-induced CO₂ pulse event or pulse event. Abiotic CO₂ pulses occur when water displaces CO₂ from soil pores^{18,19} or dissolves carbonates²⁰. Biotic CO₂ pulses, known as the Birch effect, result from microbial respiration surges following the sudden availability of labile carbon and nutrients^{13,21–24}. This resource pulse may arise from the release of intracellular solutes, microbial lysis and physical breakdown of soil aggregates^{23,25,26}. While abiotic CO₂ pulses are short lived^{20,27}, Birch-effect CO₂ pulses can last several days and lead to significant substantial ecosystem carbon losses during the day and night^{13,15,28,29}. Microbial respiration can increase 60–80-fold after rainfall¹⁵, then decay as the soil dries^{13–15,21,30}, contributing up to 40% to total soil respiration during the growing season²⁹ and 5–10%

A full list of affiliations appears at the end of the paper. ✉ e-mail: ngoc.nguyen@berkeley.edu; trevorkeenan@berkeley.edu

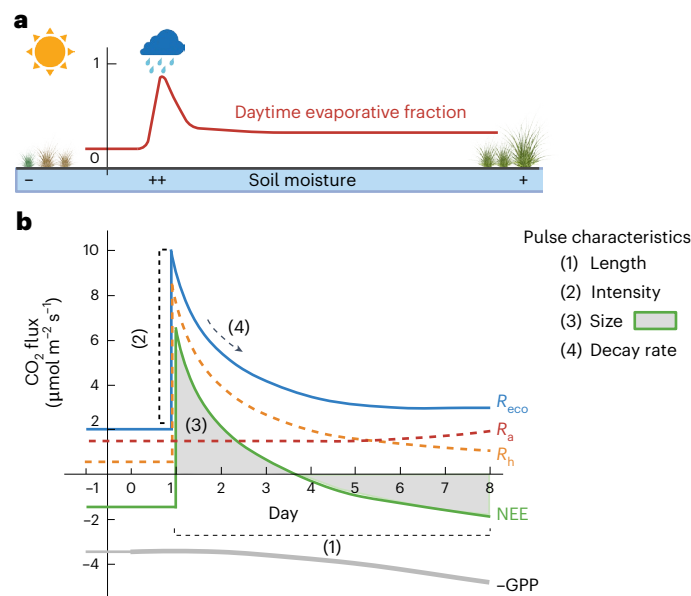


Fig. 1 | Conceptual diagram of pulse events. **a**, Pulse events, or rain-induced CO₂ pulse events, are defined as ecological processes in which rainfall on previously dry soils triggers pulses of soil CO₂ and, consequently, pulses of net CO₂ flux (NEE). The horizontal bar represents the soil moisture gradient, daytime EF and vegetation growth before and after rain pulses. **b**, The dynamics of daily ecosystem carbon fluxes (R_a , R_h , R_{eco} , NEE and GPP) during pulse events and

CO₂ pulse characteristics: (1) length, (2) intensity, (3) size and (4) decay rate (Methods). R_{eco} is ecosystem respiration, R_a is autotrophic respiration and R_h is heterotrophic respiration. The NEE of CO₂ flux was measured directly from eddy-covariance towers. We plotted $-GPP$ instead of GPP for visualization purposes. The first-principle relationship between carbon fluxes can be described as follows: $NEE = R_a + R_h - GPP$, $R_{eco} = R_a + R_h$. Figure created with [BioRender.com](https://www.biorender.com).

to annual net ecosystem productivity (NEP)²⁸ (Fig. 1). The intensity of these pulses has been linked to hydrologic dynamics such as the intensity/frequency of rewetting and antecedent soil dryness^{13–15,18}. By contrast, plant–microbe interactions are often neglected, mainly because previous research focused on soil-centric environments or vegetation-senescence periods^{12,20}. The combined effects of vegetation, soil properties, climate and hydrologic dynamics on rain-induced carbon losses, therefore, remains poorly constrained.

Modelling these pulse-driven events is particularly challenging due to the sporadic and discontinuous nature of rainfall in drylands⁶. Most models effectively incorporate steady-state processes such as responses to temperature, long-term changes in soil moisture and vegetation productivity. However they largely neglect non-steady-state but important rain-induced carbon losses^{15,31–35}. This omission could lead to substantial underestimation of soil and ecosystem respiration, potentially resulting in inaccurate estimates of related processes such as gross primary productivity (GPP)^{14,36}. The lack of a broad characterization of dryland rain-induced CO₂ pulses has hindered the integration of these processes into existing ecosystem models.

Detecting rain-induced carbon losses across diverse dryland ecosystems is a critical first step to improving model predictions of dryland carbon fluxes. FLUXNET, which provides global eddy-covariance measurements of net CO₂ flux or net ecosystem exchange (NEE)^{37,38}, offers a unique opportunity to analyse CO₂ pulse responses at high temporal resolution. Rainfall-driven ecosystem respiration (R_{eco}) pulses manifest as net CO₂ anomalies, because microbial activity typically responds more rapidly to rainfall than vegetation^{6,11,26,31,39–41} (Fig. 1). However, FLUXNET does not provide continuous measurements of soil and ecosystem respiration³⁶, making NEE the variable of choice for detecting rain-induced carbon losses. Furthermore, conventional eddy-covariance partitioning methods estimate R_{eco} using an Arrhenius-type temperature function^{34,42} (equation (2)), which may not effectively capture rain-induced carbon losses^{35,36}. NEE observations from FLUXNET thus provide a unique opportunity to both identify and characterize soil CO₂ pulses.

Here, we combine a novel manually labelled database of 1,857 pulse events with machine learning frameworks to detect rain-induced carbon losses across 34 dryland eddy-covariance sites, encompassing 323 years of half-hourly observations. Using this database, we characterize the spatiotemporal drivers of net CO₂ pulse dynamics and bias-correct R_{eco} estimates derived from the conventional night-time partitioning method. We refer to this integrated detection and correction approach as FluxPulse. Our results highlight widespread underestimation of carbon losses during rain-pulse events in global drylands, with important implications for dryland carbon balance variability under climate change.

Widespread underestimation of pulse event carbon fluxes

We found that pulse events greatly impact the dryland carbon balance, contributing $16.9\% \pm 2.8\%$ (mean \pm the margin of error) to annual ecosystem respiration (R_{eco}) and $9.6\% \pm 2.2\%$ to NEP across sites (Fig. 2a). These findings align with previous research showing that soil CO₂ pulses contribute 5–10% of annual NEP in a mid-latitude forest, with carbon losses during pulse events being comparable with the annual NEP of many ecosystems^{15,28}.

Despite this substantial contribution, we found that all four existing eddy-covariance partitioning methods significantly underestimate R_{eco} and GPP during the pulse events when compared our FluxPulse bias-corrected fluxes (Fig. 2b). Underestimation for R_{eco} ranged from $13.4\% \pm 9.4\%$ to $26.7\% \pm 9.0\%$, while GPP ranged from $17.8\% \pm 11.7\%$ to $33.8\% \pm 14.2\%$ (Fig. 2b,c). Since GPP is estimated as the difference between R_{eco} and NEE, any underestimation of R_{eco} directly affects the GPP estimates (Fig. 1).

The four partitioning methods examined include parametric approaches (night-time method³⁴ and daytime method⁴²) and machine learning methods (double machine learning (DML)⁴³ and artificial neural networks (NN)⁴⁴) (Methods). We performed a one-tailed paired *t*-test to compare partitioning methods, which revealed a clear ranking of R_{eco} underestimation during pulse events, with: NN ($13.4\% \pm 9.4\%$) < DML

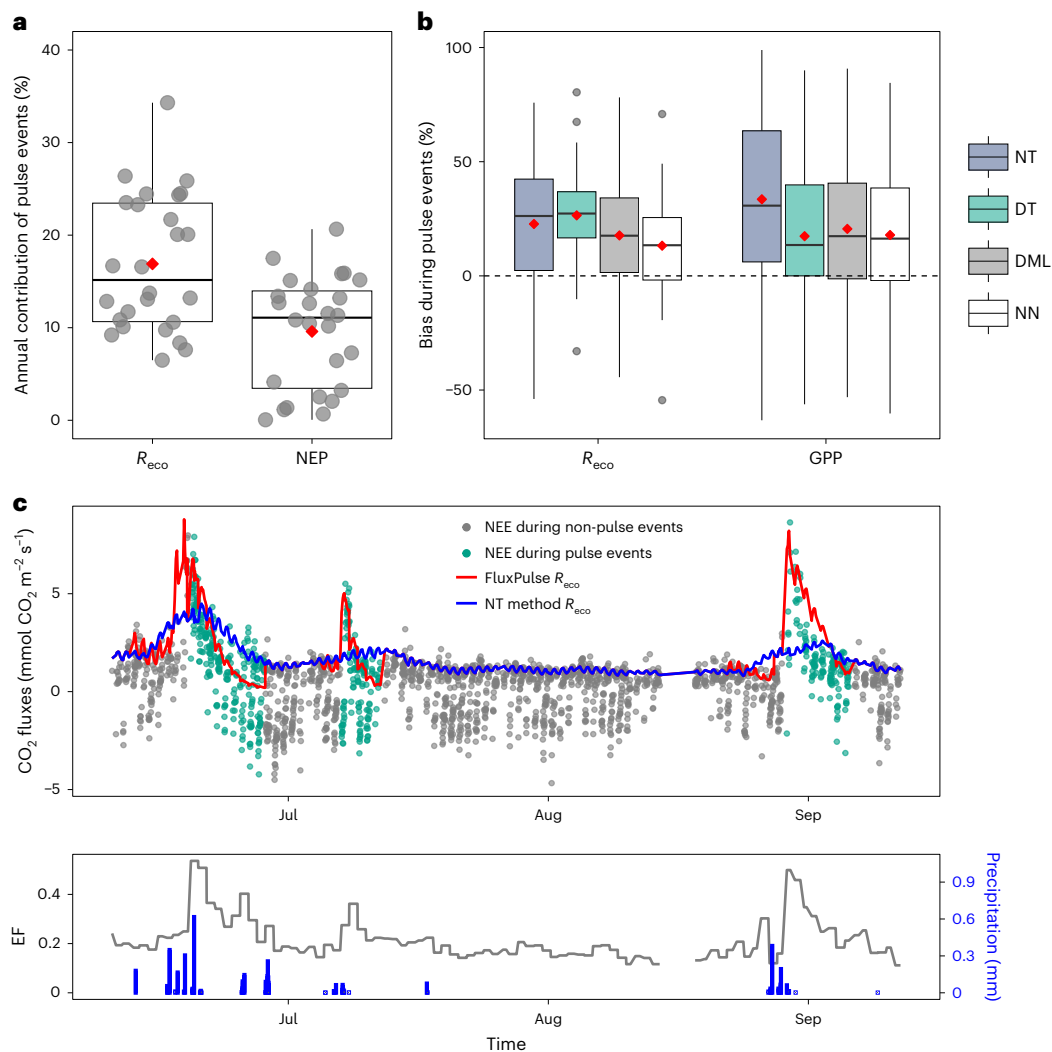


Fig. 2 | Partitioning methods underestimate pulse event CO₂ fluxes.

a, Contributions of ecosystem respiration (R_{eco}) and NEP during pulse events to annual R_{eco} and NEP. Pulse event contribution (%) at each individual site is calculated as $\frac{\text{sum}(x)}{\text{sum}(y)} \times 100$, in which x is FluxPulse bias-corrected half-hourly R_{eco} or NEP during pulse events, and y is all available half-hourly bias-corrected R_{eco} or NEP (two sites whose NEP contribution values greater than three times of standard deviation were removed in the boxplots for visualization purposes). **b**, Performance of partitioning methods on estimating ecosystem respiration (R_{eco}) and GPP during pulse events. The fluxes estimated from four methods were compared with the FluxPulse bias-corrected R_{eco} and GPP, which are the night-time (NT) method, the daytime method (DT), DML and NN (Methods). The flux underestimation is calculated as $\frac{\text{sum}(x) - \text{sum}(y)}{\text{sum}(x)} \times 100$, in which x is the FluxPulse half-hourly R_{eco} or GPP and y is the partitioned half-hourly R_{eco} or GPP from different partitioning methods. Thus, the positive y values indicate flux underestimations, while negative y values indicate flux overestimations. In **a** and

b, 28/34 sites whose FluxPulse significantly reduces biases ($P < 0.001$) (Extended Data Fig. 6) are included. The boxes represent interquartile range (IQR) marked with 25th, 50th and 75th percentile and whiskers extending to 1.5 times the IQR. The red diamonds represent the mean of the boxplots. In **b**, all the boxplots have the mean of distribution statistically greater than 0 (one-tailed paired t -test, $P < 0.01$). **c**, Half-hourly time series visualization of NEE, ecosystem respiration estimated by the FluxPulse bias-corrected algorithm (FluxPulse R_{eco}) and by the NT method R_{eco} at the Albuera site, Spain (ES-Abr) in 2017. The NEE points corresponding to pulse events are manually labelled in green, while others are labelled in black. The red line represents FluxPulse bias-corrected R_{eco} , and the blue line represents R_{eco} estimated by the NT method. In the lower graph, the black line shows the daytime EF, calculated daily. The blue bars are precipitation downscaled from ERA-Interim reanalysis data products, with half-hourly resolution. A sharp rise in EF and increased precipitation indicates the occurrence of pulse events and net CO₂ pulses.

($17.9\% \pm 9.7\%$) < night-time ($22.9\% \pm 10.7\%$), daytime method ($26.7\% \pm 9.0\%$) ($P < 0.01$) (Fig. 2b). The NN approach underestimates R_{eco} the least, followed by the DML approach, while the night-time and daytime methods exhibit similar underestimations. Regarding GPP during pulse events, the night-time method shows the largest bias ($P < 0.01$): night-time ($33.8\% \pm 14.2\%$), daytime ($17.8\% \pm 11.7\%$), DML ($20.9\% \pm 12.4\%$) and NN ($18.2\% \pm 11.2\%$) (Fig. 2b).

Among the four partitioning methods assessed, machine learning approaches (NN and DML) underestimate R_{eco} the least, highlighting their improved performance at capturing non-steady-state events compared with the parametric approaches assessed⁴⁴. The underestimation of R_{eco} also leads to underestimation of GPP, particularly in

methods where GPP is calculated as R_{eco} minus NEE. The night-time method shows the largest GPP underestimations (Fig. 2b), often leading to systematic negative half-hourly GPP during the first few days of pulse events³⁶.

Actual biases in existing methods for estimating R_{eco} and GPP during pulse events could be much higher than those reported here. Our approach assumes minimal GPP during the first day of pulse events, using maximum NEE to estimate maximum R_{eco} , given that the GPP estimates are not reliable during pulse events (Methods). Minimal GPP at the beginning of pulse events could happen in some sites due to the intensive prepulse dryness; however, we demonstrate that this is not the case in most sites (Extended Data Fig. 7). Even though

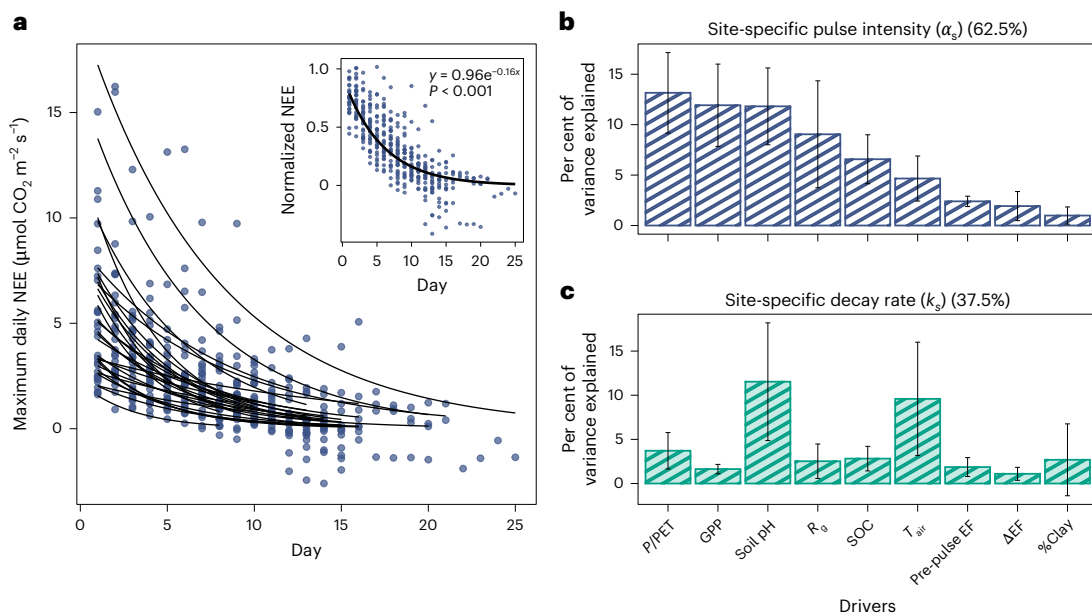


Fig. 3 | Rain-induced CO₂ pulses decay across sites. **a**, NEE declines over time during pulse events across 29/34 sites with statistically significant site-specific decay rate ($P < 0.05$). The P value is derived from the statistical test for parameter estimates when fitting the decay function to the data. The y axis of the outer graph is the maximum of daily mean NEE from all pulse events from one site, and the x axis is the number of days after the pulse event starts, assuming that rainfall starts on day 1. The first-order kinetic reaction function is fitted for each site and displayed as black lines (equation (2)). The inset presents the same data but normalized by site-specific pulse intensity (α_s) and reveals a convergent decay rate across sites ($k_s = 0.16 \pm 0.01$, $P < 0.001$). **b, c**, A relative importance analysis of nine predictors for site-specific pulse intensity (α_s) (**b**) and decay rate (k_s) (**c**)

the underestimates we quantified are conservative, our FluxPulse bias-correction method significantly reduces biases compared with existing partitioning methods, providing more reliable estimates of carbon fluxes during pulse events.

Convergence in decay rate of rain-induced carbon pulses

Analysing more than 1,800 pulse events, we found that rain-induced carbon pulses decay exponentially after rain events in 85% of sites, following a universal exponential decay pattern (Fig. 3a). The intensity and decay rates of these pulses are primarily governed by environmental factors related to vegetation growth, climate and soil properties (Fig. 3b,c). We calculated site-specific pulse intensity (α_s) and decay rate (k_s) by fitting a first-order kinetic reaction function to the daily maximum NEE across all pulse events at each site (Fig. 3a). The site-specific pulse intensity represents the maximum magnitude of rain-induced carbon losses at a site, and the site-specific decay rate represents how fast the carbon losses decay over time (equation (1)).

When normalizing the NEE rain-pulse response curves by site-specific pulse intensity, the response curves across sites converge to a universal decaying function ($y = 0.96 \times e^{-0.16x}$) (Fig. 3a). This indicates a consistent response of soil microbes to rain pulses, probably due to the strong relationship between soil microbes, soil water potential and evaporation⁴⁵. Prior studies have shown that cumulative soil evaporation follows a square root of time as the soil dries⁴⁶, further supporting this convergence. Our finding of the convergent decay rate could facilitate the future incorporation of the rain-pulse effect into large-scale ecosystem models. The decaying pattern of net CO₂ flux over time probably results from resource depletion (labile carbon, nitrogen, water availability and so on), changes in microbial community

composition²⁶ and vegetation growth triggered by pulse events which off-sets CO₂ release. The widespread decay pattern highlights the biological influence of rainfall on ecosystem processes beyond abiotic influences, reinforcing the importance of the Birch effect in dryland carbon cycling. Our findings align with previous regional and laboratory studies^{3,15,21,47}, confirming the widespread occurrence of the Birch effect in global drylands.

Multiple environmental factors influence the spatial dynamics of net carbon fluxes during pulse events. The three strongest predictors for the site-specific pulse intensity (α_s) across sites are aridity index (P/PET), GPP and soil pH, together with other factors explaining 62.5% variance (Fig. 3b). α_s increases with Aridity index (P/PET) and GPP but decreases with soil pH (Extended Data Fig. 9). For the site-specific decay rate (k_s), soil pH and air temperature were important predictors, with k_s increasing with soil pH and temperature (Extended Data Fig. 8). However, compared with α_s , environmental factors explain only 37.5% of variance in k_s (Fig. 3b,c). Regarding hydrologic factors, we used evaporative fraction (EF), the fraction of available energy allocated to evapotranspiration, as a proxy of soil water availability and rainfall occurrence (Methods). Despite exhibiting a strong temporal correlation with pulse intensity and size, prepulse EF (the 14-day average EF before a pulse event) and Δ EF (a measure of the rewetting intensity) do not significantly explain the spatial variability of net CO₂ pulses during pulse events (Figs. 3b,c and 4c and Extended Data Fig. 4).

Our results show that vegetation productivity is key to initial rewetting responses, as more productive sites supply more labile carbon through root exudation and plant litter, enhancing microbial decomposition^{12,15,48} (Fig. 3b,c). This also explains why soil organic carbon (SOC) shows little explanatory power for the spatial distribution of pulse characteristics (Fig. 3b,c), since SOC composition (for

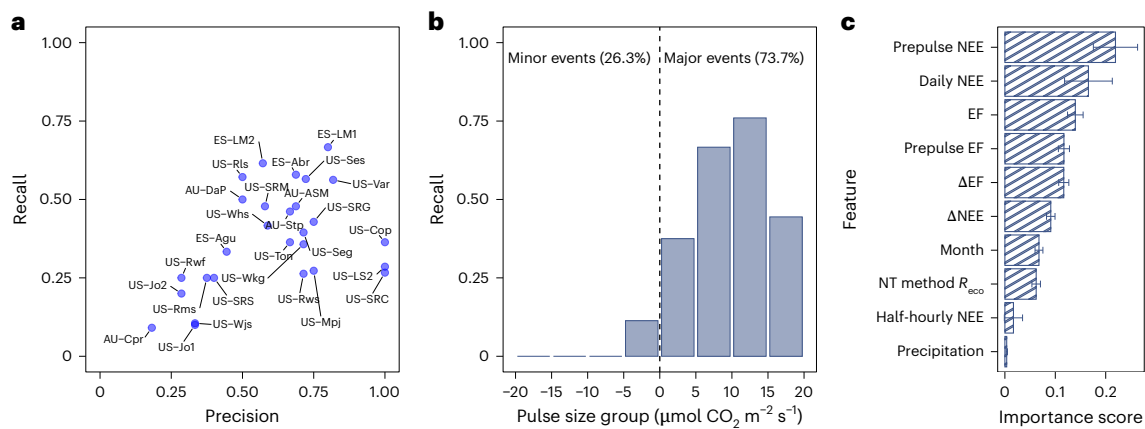


Fig. 4 | Random forest performance for detecting pulse events.

a, A performance for detecting the start date of pulse events (within ± 2 days error) across 28 qualified sites (see Supplementary Table 1 for more details). Precision is the fraction of pulse events the model detected that were real, while recall is the fraction of actual pulse events that were successfully detected by the model. **b**, The recall rate for different pulse size groups (divided into equal intervals of $5 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$, ranging from -20 to $20 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$). The

dashed line $x = 0$ defines major ($x > 0$) and minor ($x < 0$) pulse events based on the size of net CO_2 pulses. Major pulse events (73.7% of the testing set) are defined as carbon sources, while minor pulse events (26.3%) are carbon sinks. **c**, Feature importance score for all variables used in the random forest model to detect pulse events across 28 sites. The data are presented as the mean value \pm standard deviation of accumulation of the impurity decrease for each feature within each decision tree.

example, labile versus recalcitrant SOC) is likely more important than total SOC in sustaining the Birch effect^{23,49} (Fig. 3b,c). Besides GPP, the aridity index (P/PET) is a major predictor of pulse intensity across sites (Fig. 3b). Among all dry sites, wetter sites have higher pulse intensity (Extended Data Fig. 9). This suggests that wetter dryland sites have sufficient moisture to enhance vegetation growth, supplying substrates for microbial decomposition, while they are still dry enough for the Birch effect to occur.

Among all environmental factors, soil pH is a top predictor of both site-specific pulse intensity and decay rate (Fig. 3b,c). Previous studies have highlighted the influence of soil pH on microbial composition^{50–56}, with soil microbial diversity peaks at near-neutral soil pH at a continental scale⁵³. In our study, we also observe that the site-specific pulse intensity peaks at pH 6 and gradually declines as the pH increases to 8 (Extended Data Fig. 9). This could be attributed to various soil characteristics associated with soil pH, as well as physiological constraints on soil bacteria under extremely acidic or alkaline conditions⁵³. Our validation of SoilGrids pH against site-specific measurements confirms soil pH measurement robustness (Extended Data Fig. 3). Our study shows that soil pH is an important spatial driver of rain-induced carbon pulses, potentially due to its influence on soil microbial dynamics.

Detecting and characterizing rain-induced carbon pulses

Rain-induced CO_2 pulse events are non-steady-state ecological processes, making them challenging to detect and model across sites. Nevertheless, automated pulse detection is now possible, given the large amount of data and a better understanding of the complexity of rain-pulse responses. We developed an automated detection framework using a random forest binary classification algorithm, in addition to manually labelling pulse events (Methods). Our results demonstrate that although performance varies across sites, which is possibly due to data quality and site-specific measurement uncertainties, such automated detection is applicable across diverse ecosystems. We successfully detected at least 60% of pulse events in 15/28 sites (Fig. 4a). Detection is most effective for major pulse events—those with net positive CO_2 emissions, with 76% of pulses whose size between 10 and $15 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ being correctly detected across sites (Fig. 4b).

We used two common classification metrics to evaluate model performance: precision and recall. Precision is the fraction of detected pulse events that are actual pulse events, while recall is the fraction

of actual pulse events correctly detected by the model. Actual pulse events, in this case, are manually labelled pulse events. For example, if 12 pulse events occur, and the model detects 10 events, with only 7 of them being real, then precision is 7/10 (70%) and recall is 7/12 (58%) (Methods). Across sites, the model achieves a precision greater than 0.60 at 15/28 sites (Fig. 4a). Recall rates are generally lower than precision, with the highest recall rate (0.67) observed at the ES-LM1 site (Fig. 4a and Supplementary Table 1). Recall improves with the net CO_2 pulse size, reaching 0.76 for pulses ranging from 10 to $15 \mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$ (Fig. 4b). Although the model did not effectively capture pulse events with negative pulse sizes, it performed well on most pulse events (73.7% of the pulse events in the testing set), particularly those with positive pulse sizes (Fig. 4b).

Among the features used for classification (Methods), prepulse NEE and daily mean NEE are the two most important predictors of the forest detection algorithm (Fig. 4c), meaning that ecosystem status before pulse events is crucial to detecting them. Pulse occurrence is more likely with positive prepulse NEE and less likely with negative prepulse NEE values when vegetation is active (Extended Data Fig. 10). In a woody savanna, the magnitude of CO_2 pulses has also been shown to be inversely correlated with prepulse R_{eco} (ref. 15).

Moreover, EF-derived indices (EF, antecedent water availability (prepulse EF) and rewetting intensity (ΔEF)) accurately describe temporal variability in pulse characteristics and improve machine learning detection performance (Figs. 3b,c and 4c and Extended Data Fig. 4). Lower prepulse EF and higher EF associate with a higher chance of pulse events (Extended Data Fig. 10). Furthermore, our analysis across multiple sites indicates that more intense rewetting events (ΔEF) and drier prepulse periods (prepulse EF) are linked with higher pulse intensity (ΔNEE) and greater pulse size (Extended Data Fig. 4). This indicates that larger pulses are more likely after prolonged dry periods followed by intense rewetting. Our findings suggest that under future scenarios of more intense precipitation and prolonged dry intervals, rain-induced carbon losses will probably intensify⁵⁷.

Including site-static variables (for example, soil pH, vegetation productivity and aridity index) does not enhance model performance, though adding month as a temporal variable enhances results (Fig. 4c). Interestingly, precipitation is not a substantial predictor of pulse occurrence or intensity (Fig. 4c and Extended Data Fig. 4), most likely because EF already captures rainfall effects, and not all rainfall events trigger CO_2 pulses.

Overall, EF is an integrated index that captures water dynamics, including the wetting intensity and antecedent water availability. Compared with precipitation, EF-derived indices play an important role in the pulse detection algorithm (Fig. 4c) and they significantly correlate with pulse intensity and size across most sites (Extended Data Fig. 4). While precipitation is undoubtedly a major driver of rain-induced CO₂ pulses, its direct influence is difficult to detect due to under-report rainfall from using the tipping bucket method and landscape heterogeneity^{58,59}. EF, which integrates both wetting intensity and antecedent water availability, could thus be a promising index for water availability particularly at eddy-covariance sites where precipitation and soil moisture data are limited.

While the start date of pulse events is typically well-defined, determining the end date is more ambiguous, as the end date corresponds to the system either returning to prepulse conditions or reaching a stabilized state. However, since carbon emissions are greatest in the first few days of pulse events, potential misidentification of the end date is unlikely to impact overall CO₂ loss estimates. Future work should further optimize detection algorithms and test whether rain-induced CO₂ pulses occur in other biomes whose high vegetation activity could mask their signals. Overall, the random forest gives insight into the drivers of pulse responses, enabling a deeper understanding of pulse responses in dryland ecosystems.

Implications for terrestrial carbon cycle projections

Drylands drive the interannual variability and trend of global land carbon sink, but the processes underlying the rain-induced carbon losses responsible for carbon flux variability have not been widely incorporated in terrestrial ecosystem carbon models⁶⁰. Our study demonstrates that rain-induced carbon losses follow a significant and consistent decay pattern across sites and contribute significantly to the carbon balance of drylands (16.9% ± 2.8% of annual ecosystem respiration and 9.6% ± 2.2% of annual NEP). We show that existing partitioning methods substantially underestimate carbon fluxes in drylands, potentially affecting higher-level ecosystem carbon models and upscaling studies that rely on flux partitioned data. We resolve persistent biases in current eddy-covariance partitioning methods by providing FluxPulse, a bias-corrected dataset of ecosystem carbon fluxes in drylands. Our results highlight the importance of effectively incorporating rain-induced carbon losses from pulse events into models from site to global scales. This is particularly important given the projected expansion of drylands in the twenty-first century⁶¹ and rapid changes in global precipitation patterns favouring more intense and less frequent rainfall⁵⁷.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41561-025-01754-9>.

References

- Reynolds, J. F. et al. Global desertification: building a science for dryland development. *Science* **316**, 847–851 (2007).
- Whitford, W. G. & Duval, B. D. In *Ecology of Desert Systems* 2nd edn (eds Whitford, W. G. & Duval, B. D.) 47–72 (Academic Press, 2020); <https://doi.org/10.1016/B978-0-12-815055-9.00003-5>
- Ahlström, A. et al. The dominant role of semi-arid ecosystems in the trend and variability of the land CO₂ sink. *Science* **348**, 895–899 (2015).
- Poulter, B. et al. Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle. *Nature* **509**, 600–603 (2014).
- Sitch, S. et al. Trends and drivers of terrestrial sources and sinks of carbon dioxide: an overview of the TRENDY project. *Global Biogeochem. Cycles* **38**, e2024GB008102 (2024).
- Noy-Meir, I. Desert ecosystems: environment and producers. *Ann. Rev. Ecol. System.* **4**, 25–51 (1973).
- Feldman, A. F. et al. Moisture pulse-reserve in the soil-plant continuum observed across biomes. *Nat. Plants* **4**, 1026–1033 (2018).
- Sala, O. E., Gherardi, L. A., Reichmann, L., Jobbágy, E. & Peters, D. Legacies of precipitation fluctuations on primary production: theory and data synthesis. *Philos. Trans. R. Soc. Lond. B* **367**, 3135–3144 (2012).
- Kannenbergh, S. A. et al. Quantifying the drivers of ecosystem fluxes and water potential across the soil-plant-atmosphere continuum in an arid woodland. *Agric. For. Meteorol.* **329**, 109269 (2023).
- Gherardi, L. A. & Sala, O. E. Effect of interannual precipitation variability on dryland productivity: a global synthesis. *Global Change Biol.* **25**, 269–276 (2019).
- Huxman, T. E. et al. Response of net ecosystem gas exchange to a simulated precipitation pulse in a semi-arid grassland: the role of native versus non-native grasses and soil texture. *Oecologia* **141**, 295–305 (2004).
- Barnard, R. L. Rewetting of soil: revisiting the origin of soil CO₂ emissions. *Soil Biol. Biochem.* **4**, 107819 (2020).
- Jarvis, P. et al. Drying and wetting of Mediterranean soils stimulates decomposition and carbon dioxide emission: the ‘Birch effect’. *Tree Physiol.* **27**, 929–940 (2007).
- Rousk, J. & Brangari, A. C. Do the respiration pulses induced by drying-rewetting matter for the soil-atmosphere carbon balance? *Glob. Chang. Biol.* **28**, 3486–3488 (2022).
- Xu, L., Baldocchi, D. D. & Tang, J. How soil moisture, rain pulses, and growth alter the response of ecosystem respiration to temperature. *Global Biogeochem. Cycles* **18**, GB4002 (2004).
- MacBean, N. et al. Dynamic global vegetation models underestimate net CO₂ flux mean and inter-annual variability in dryland ecosystems. *Environ. Res. Lett.* **16**, 094023 (2021).
- Smith, W. K. et al. Remote sensing of dryland ecosystem structure and function: progress, challenges, and opportunities. *Remote Sens. Environ.* **233**, 111401 (2019).
- Huxman, T. E. et al. Precipitation pulses and carbon fluxes in semiarid and arid ecosystems. *Oecologia* **141**, 254–268 (2004).
- Maier, M., Schack-Kirchner, H., Hildebrand, E. E. & Holst, J. Pore-space CO₂ dynamics in a deep, well-aerated soil. *Eur. J. Soil Sci.* **61**, 877–887 (2010).
- Inglisma, I. et al. Precipitation pulses enhance respiration of Mediterranean ecosystems: the balance between organic and inorganic components of increased soil CO₂ efflux. *Global Change Biol.* **15**, 1289–1301 (2009).
- Birch, H. F. The effect of soil drying on humus decomposition and nitrogen availability. *Plant Soil* **10**, 9–31 (1958).
- Bottner, P. Response of microbial biomass to alternate moist and dry conditions in a soil incubated with ¹⁴C- and ¹⁵N-labelled plant material. *Soil Biol. Biochem.* **17**, 329–337 (1985).
- Fierer, N. & Schimel, J. P. A proposed mechanism for the pulse in carbon dioxide production commonly observed following the rapid rewetting of a dry soil. *Soil Sci. Soc. Am. J.* **67**, 798–805 (2003).
- Schimel, J. P. Life in dry soils: effects of drought on soil microbial communities and processes. *Annu. Rev. Ecol. Evol. Syst.* **49**, 409–432 (2018).
- Rutledge, S., Campbell, D. I., Baldocchi, D. & Schipper, L. A. Photodegradation leads to increased carbon dioxide losses from terrestrial organic matter. *Global Change Biol.* **16**, 3065–3074 (2010).

26. Placella, S. A., Brodie, E. L. & Firestone, M. K. Rainfall-induced carbon dioxide pulses result from sequential resuscitation of phylogenetically clustered microbial groups. *Proc. Natl Acad. Sci. USA* **109**, 10931–10936 (2012).
27. Serrano-Ortiz, P. et al. Hidden, abiotic CO₂ flows and gaseous reservoirs in the terrestrial carbon cycle: review and perspectives. *Agric. For. Meteorol.* **150**, 321–329 (2010).
28. Lee, X., Wu, H.-J., Sigler, J., Oishi, C. & Siccama, T. Rapid and transient response of soil respiration to rain. *Global Change Biol.* **10**, 1017–1026 (2004).
29. Yan, L., Chen, S., Xia, J. & Luo, Y. Precipitation regime shift enhanced the rain pulse effect on soil respiration in a semi-arid steppe. *PLoS ONE* **9**, e104217 (2014).
30. Liu, X., Wan, S., Su, B., Hui, D. & Luo, Y. Response of soil CO₂ efflux to water manipulation in a tallgrass prairie ecosystem. *Plant Soil* **240**, 213–223 (2002).
31. Migliavacca, M. et al. Semiempirical modeling of abiotic and biotic factors controlling ecosystem respiration across eddy covariance sites. *Global Change Biol.* **17**, 390–409 (2011).
32. Raich, J. W. & Potter, C. S. Global patterns of carbon dioxide emissions from soils. *Global Biogeochem. Cycles* **9**, 23–36 (1995).
33. Reichstein, M. et al. Modeling temporal and large-scale spatial variability of soil respiration from soil water availability, temperature and vegetation productivity indices. *Global Biogeochem. Cycles* **17**, GB002035 (2003).
34. Reichstein, M. et al. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biol.* **11**, 1424–1439 (2005).
35. Xu, L. & Baldocchi, D. D. Seasonal variation in carbon dioxide exchange over a Mediterranean annual grassland in California. *Agric. For. Meteorol.* **123**, 79–96 (2004).
36. Jung, M. et al. Technical note: flagging inconsistencies in flux tower data. *Biogeosciences* **21**, 1827–1846 (2024).
37. Baldocchi, D. et al. FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull. Am. Meteorol. Soc.* **82**, 2415–2434 (2001).
38. Baldocchi, D. D. How eddy covariance flux measurements have contributed to our understanding of Global Change Biology. *Global Change Biol.* **26**, 242–260 (2020).
39. Chen, S., Lin, G., Huang, J. & Jenerette, G. D. Dependence of carbon sequestration on the differential responses of ecosystem photosynthesis and respiration to rain pulses in a semiarid steppe. *Global Change Biol.* **15**, 2450–2461 (2009).
40. Huang, B. & Nobel, P. S. Hydraulic conductivity and anatomy along lateral roots of cacti: changes with soil water status. *New Phytol.* **123**, 499–507 (1993).
41. Roby, M. C., Scott, R. L. & Moore, D. J. P. High vapor pressure deficit decreases the productivity and water use efficiency of rain-induced pulses in semiarid ecosystems. *JGR Biogeosci.* **125**, e2020JG005665 (2020).
42. Lasslop, G. et al. Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. *Global Change Biol.* **16**, 187–208 (2010).
43. Cohrs, K. -H., Varando, G., Carvalhais, N., Reichstein, M. & Camps-Valls, G. Causal hybrid modeling with double machine learning—applications in carbon flux modeling. *Mach. Learn. Sci. Technol.* <https://doi.org/10.1088/2632-2153/ad5a60> (2024).
44. Tramontana, G. et al. Partitioning net carbon dioxide fluxes into photosynthesis and respiration using neural networks. *Global Change Biol.* **26**, 5235–5253 (2020).
45. Orchard, V. A. & Cook, F. J. Relationship between soil respiration and soil moisture. *Soil Biol. Biochem.* **15**, 447–453 (1983).
46. Gardner, W. R. Solutions of the flow equation for the drying of soils and other porous media. *Soil Sci. Soc. Am. J.* **23**, 183–187 (1959).
47. Jenerette, G. D., Scott, R. L. & Huxman, T. E. Whole ecosystem metabolic pulses following precipitation events. *Funct. Ecol.* **22**, 924–930 (2008).
48. Tang, J., Baldocchi, D. D. & Xu, L. Tree photosynthesis modulates soil respiration on a diurnal time scale. *Global Change Biol.* **11**, 1298–1304 (2005).
49. Deng, Q., Hui, D., Chu, G., Han, X. & Zhang, Q. Rain-induced changes in soil CO₂ flux and microbial community composition in a tropical forest of China. *Sci. Rep.* **7**, 5539 (2017).
50. Chapin, F. S., Matson, P. A. & Vitousek, P. M. *Principles of Terrestrial Ecosystem Ecology* (Springer, 2011); <https://doi.org/10.1007/978-1-4419-9504-9>
51. Dacal, M., Bradford, M. A., Plaza, C., Maestre, F. T. & García-Palacios, P. Soil microbial respiration adapts to ambient temperature in global drylands. *Nat. Ecol. Evol.* **3**, 232–238 (2019).
52. Fierer, N., Grandy, A. S., Six, J. & Paul, E. A. Searching for unifying principles in soil ecology. *Soil Biol. Biochem.* **41**, 2249–2256 (2009).
53. Lauber, C. L., Hamady, M., Knight, R. & Fierer, N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* **75**, 5111–5120 (2009).
54. Schneider, T. et al. Who is who in litter decomposition? Metaproteomics reveals major microbial players and their biogeochemical functions. *ISME J* **6**, 1749–1762 (2012).
55. Sinsabaugh, R. L. et al. Stoichiometry of soil enzyme activity at global scale. *Ecol. Lett.* **11**, 1252–1264 (2008).
56. Tripathi, B. M. et al. Soil pH mediates the balance between stochastic and deterministic assembly of bacteria. *ISME J.* **12**, 1072–1083 (2018).
57. Feldman, A. F. et al. Plant responses to changing rainfall frequency and intensity. *Nat. Rev. Earth Environ.* **5**, 276–294 (2024).
58. Gebler, S. Actual evapotranspiration and precipitation measured by lysimeters: a comparison with eddy covariance and tipping bucket. *Hydrol. Earth Syst. Sci.* **19**, 2145–2161 (2015).
59. Legates, D. R. & DeLiberty, T. L. Precipitation measurement biases in the United States. *J. Am. Water Res. Assoc.* **29**, 855–861 (1993).
60. Metz, E.-M. et al. Soil respiration-driven CO₂ pulses dominate Australia's flux variability. *Science* **379**, 1332–1335 (2023).
61. Huang, J., Yu, H., Guan, X., Wang, G. & Guo, R. Accelerated dryland expansion under climate change. *Nat. Clim. Change* **6**, 166–171 (2016).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

¹Department of Environmental Science, Policy, and Management, University of California, Berkeley, Berkeley, CA, USA. ²European Commission, Joint Research Centre, Ispra, Italy. ³Department of Environmental Science, University of Arizona, Tucson, AZ, USA. ⁴National Research Council, Institute of Research on Terrestrial Ecosystems, Monterotondo Scalo, Italy. ⁵University of Tuscia, DIBAF, Viterbo, Italy. ⁶Max Planck Institute of Biogeochemistry, Department of Biogeochemical Integration, Jena, Germany. ⁷Universitat de València, València, Spain. ⁸Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, USA. ⁹H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA. ¹⁰Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA. ¹¹Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. ✉e-mail: ngoc.nguyen@berkeley.edu; trevorkeenan@berkeley.edu

Methods

Data sources

We collected openly available eddy-covariance data from FLUXNET2015⁶², AmeriFlux (www.ameriflux.lbl.gov) and ICOS (Integrated Carbon Observation System) Warm Winter 2020 (ref. 63) (www.icos-cp.eu), all produced through ONEFlux processing under a CC-BY-4.0 license⁶². We used only original half-hourly measurements extracted from the gapfilled versions of NEE of CO₂ (NEE_VUT_REF), latent heat flux (LE_F_MDS) and sensible heat flux (H_F_MDS), along with meteorological variables such as air temperature (TA_F_MDS) and incoming shortwave radiation (SW_IN_F_MDS). For precipitation, we gathered both tower-measured precipitation (*P*) and downscaled precipitation (P_ERA) from ERA-Interim reanalysis data products published via the above eddy-covariance networks⁶⁴.

We selected eddy-covariance sites that satisfied the following criteria: (1) ratio of annual precipitation to potential evapotranspiration (*P*/PET) < 0.65, a common definition of a dryland⁶⁵, (2) at least 4 years of data and (3) short and sparse vegetation (typically height < 2 m and area < 30% of tree cover) with IGBP vegetation types belonging to grasslands, savannas, open shrublands, closed shrublands and woody savannas (WSA). We selected ecosystems with short and sparse vegetation to minimize rain loss from vegetation interception and reduce strong vegetation interference with rain-induced CO₂ pulse signals detected in NEE measurements. In total, our data contained 34 sites across North America, Europe and Australia, with mean annual precipitation (MAP) ranging from 245 to 1,180 mm, mean annual temperature ranging from 7 °C to 29 °C and record length ranging from 4 to 21 years in each site (Extended Data Fig. 1 and Supplementary Table 1).

Identification of pulse events

We established a foundational database of pulse events by manually labelling the start and end dates of these events, across 323 site-years of data in 34 sites. We then developed a machine learning algorithm to automatically detect pulse events from half-hourly eddy-covariance data, which was trained on the manually labelled dataset.

Pulse event indicator. Rainfall is typically measured by the tipping bucket method in the eddy-covariance network; nevertheless, this method is known to under-report rainfall due to instrument limitations and landscape heterogeneity^{58,59}. We confirmed that many tower data probably under-reported precipitation by noting that precipitation downscaled from ERA (P_ERA) is more synchronous with net CO₂ pulses than the in situ-measured precipitation (*P*) (Extended Data Fig. 2). We therefore used P_ERA as the primary rainfall record.

In addition to rainfall, we used daytime EF as an index of soil water availability and rainfall occurrence. EF is the fraction of available energy allocated to evapotranspiration from the land surface, which we calculated on a daily scale as the average daytime ratio between latent heat flux and the sum of latent and sensible heat fluxes⁶⁶. Daytime EF varies between 0 and 1, with 1 indicating a non-water-limiting condition as all available energy is converted to latent heat and 0 indicating a water-limited condition since all available energy is converted to sensible heat. EF range is 0.19 ± 0.18 to 0.52 ± 0.15 across 34 studied sites, including both growing and non-growing seasons. Rainfall after a long dry period creates a surge of EF, and dry ecosystems rarely get too wet for EF to saturate, making EF a potential indicator of soil water availability and rainfall occurrence. Compared with in situ soil water content measurements, EF is broadly available and can be calculated consistently across eddy-covariance sites.

Manual labelling of pulse events. We manually labelled the start and end dates of pulse events via observations of net CO₂ pulses, aiming to study their spatiotemporal dynamics and to develop supervised machine learning models for labelling these events (Extended Data Fig. 10 for detailed criteria for labelling). We used original half-hourly

time series data on NEE of CO₂, daytime EF and downscaled precipitation (P_ERA) from ERA-Interim reanalysis data. Across 34 sites over 323 site-years, we manually labelled 1857 pulse events, with durations ranging from 2 to 26 days. The start date of the pulse event was labelled based on satisfying the following conditions: (1) NEE suddenly increases sharply, (2) NEE gradually decays over at least 2 days after the peak day to distinguish the start of pulses from random noises, (3) EF suddenly increases sharply and (4) positive precipitation (Supplementary Table 2). Sometimes, condition (4) was not satisfied, but we still marked the start of the pulse event because it is likely that rainfall was under-reported. The end date of the pulse event was defined as the date when night-time NEE stabilizes and approximates its prepulse magnitude (Extended Data Fig. 10).

Machine learning labelling of pulse events. We used random forests to automatically label pulse events at a half-hourly scale, since random forests or tree-based models in general, have been shown to outperform deep learning on tabular datasets⁶⁷. Furthermore, random forest can inherently handle missing time series data and is more time-efficient compared with deep learning algorithms. Specifically, we applied the random forest binary classification algorithm using the function 'RandomForestClassifier' in the Python 'sklearn' package. The algorithm, trained on previously manually labelled datasets, processes FLUXNET formatted data to label measurements as pulse events or not on a half-hourly scale. To train and evaluate this model, we first selected 28 out of 34 sites that each has at least 5 years of data and 20 recorded pulse events. We trained one foundation model using data from all selected sites, using the last 3 years at each site as the testing set; the rest of the observations were randomly split into a training set and a validation set of ratio 80%:20%. This approach not only enhances the generalization capability of our model but also mitigates the risk of overfitting. The validation set was used for model selection purposes (such as tuning hyper-parameters), while the test set was used to evaluate the model's final performance. The features used by the random forest model include hydrologic and carbon flux characteristics that are important signals for the pulse events, including rewetting intensity (ΔEF), antecedent water availability (prepulse EF), pulse intensity (ΔNEE) and antecedent ecosystem productivity (prepulse NEE) (Methods). The full list can be found in Fig. 4c.

We evaluate this model, again with respect to the test set, using two common metrics for classification tasks: precision and recall. Precision is the fraction of pulse events the model detected that were actually real ($\frac{\text{true positive}}{\text{true positive} + \text{false positive}}$), while recall is the fraction of actual pulse events that were successfully detected by the model ($\frac{\text{true positive}}{\text{true positive} + \text{false negative}}$). For example, if there are 12 pulse events in total, and the model detected 10 events with only 7 of them being actual pulse events, then precision is 7/10 and recall is 7/12.

Finally, the random forest model allowed us to investigate feature importance, that is, how relatively useful each covariate is in predicting the outcome. In particular, for each feature, we computed the normalized total reduction in the Gini impurity across decision tree leaves. Higher Gini importance indicates that the feature reduces more uncertainty in prediction on average and hence is more important⁶⁸. We report these scores in Fig. 4c.

Characterizing rain-induced carbon pulses

Length, intensity and size. We characterized rain-induced carbon pulses within the eddy-covariance network by defining the length, intensity and size of NEE during each event (equation (1) and Fig. 1). Even though the decay rate is an important feature of pulse events (Fig. 1), due to the uncertainties of half-hourly data and heterogeneity in data quality across sites, it is not possible to derive the decay rate for each event. Instead, we characterized the site-specific decay rate

for all pulse events within one site (see ‘Site-specific pulse intensity and decay rate’). Regarding pulse length, we defined it as the number of days of the pulse events from the start to the end dates, which was directly calculated from the manually labelled dataset. Pulse intensity is, conceptually, the initial response of microbial respiration (R_h) to rainfall, which was calculated as the change in NEE (ΔNEE) between the mean of 2 days after and the mean of 2 days before rainfall (Fig. 1). NEE was used instead of R_h to calculate the pulse intensity since vegetation has a lagged response to rainfall, hence ΔNEE is equal to ΔR_h on the first few days of the pulse events, and the eddy-covariance network does not provide direct measurements of R_h (Fig. 1). Regarding pulse size, we defined it as the sum of the daily mean of half-hourly NEE during the whole period of pulse events (Fig. 1).

Site-specific pulse intensity and decay rate. To estimate site-specific pulse intensity and decay rate, we fitted a first-order kinetic equation, which aims to mimic the litter decomposition function⁵⁰

$$\text{NEE}_{i,s} = \alpha_s \times e^{-k_s \times i}, \quad (1)$$

where $\text{NEE}_{i,s}$ is the maximum of daily mean NEE on day i from all pulse events in site s (equation (1)). α_s and k_s are the site-specific pulse intensity and decaying rate of site s , respectively (equation (1)). The site-specific pulse intensity α_s is obtained by fitting equation (1) to all pulse events from one site, which is different from the pulse intensity of individual pulse events mentioned above (ΔNEE), even though they represent a similar concept of the initial response of the net carbon flux to rainfall.

Existing R_{eco} /GPP partitioning methods

Parametric methods. The partitioning methods we utilize partition R_{eco} and GPP from NEE based on the relationship $\text{NEE} = R_{\text{eco}} - \text{GPP}$, in which NEE is the NEE of CO_2 fluxes measured directly from eddy-covariance towers, R_{eco} is ecosystem respiration and GPP is gross primary productivity (Fig. 1). We evaluated the performance of partitioning methods on estimating R_{eco} and GPP during pulse events and annually. The common parametric approaches that have been widely used by the flux community are the night-time method, which derives R_{eco} from night-time data as an Arrhenius-type function of temperature^{34,69}, and the daytime approach, which models GPP and R_{eco} based on the light-response curve of GPP and temperature response of R_{eco} (ref. 42). Both approaches model R_{eco} at half-hour j as an exponential function of air temperature (equation (2))⁶⁹

$$\text{Reco}_{\text{NTmethod},j} = R_{\text{ref}} \times e^{E_0 \times \left(\frac{1}{T_{\text{ref}} - T_0} - \frac{1}{T_{\text{air},j} - T_0} \right)}, \quad (2)$$

where $\text{Reco}_{\text{NTmethod},j}$ is R_{eco} modelled by the night-time method at half-hour j , $T_{\text{air},j}$ is the air temperature at half-hour j , R_{ref} is reference respiration at reference temperature ($T_{\text{ref}} = 288.15 \text{ K}$) and $T_0 = 227.13 \text{ K}$ and E_0 is temperature sensitivity (Fig. 2c).

Machine learning methods. In addition to evaluating the parametric approaches, we assessed machine learning methods to partition R_{eco} and GPP during pulse events, which are a hybrid model based on artificial NN⁴⁴ and a causal hybrid model based on DML⁴³. The NN approach integrates physical constraints to simultaneously estimate R_{eco} and GPP through two sub-networks and showed improved flux estimates during pulse events at the US-SRG site⁴⁴. The causal hybrid model enhances the hybrid model by incorporating principled causal knowledge and employs the DML approach for estimating causal effects⁴³. Each model incorporates variables such as soil water content, shortwave radiation, day of the year, wind speed, vapour pressure deficit and air temperature as predictors. We reimplemented both approaches and ran them on our selected sites.

FluxPulse, modelling rain-induced CO_2 pulses from flux data

We developed the FluxPulse algorithm to model rain-induced carbon pulses from eddy-covariance flux data in 34 studied global dryland sites. FluxPulse bias-corrects R_{eco} and GPP from the night-time method during pulse events and prepulse periods, in which the night-time method fails to capture the temporal dynamics (Fig. 2c). For non-rerewetting and non-prepulse periods, FluxPulse retains R_{eco} and GPP estimates from the NT method. A visualization tool for FluxPulse bias-corrected datasets can be accessed via ref. 70.

Bias-correct carbon fluxes during pulse events. To bias-correct both R_{eco} and GPP during pulse events, we first bias-corrected R_{eco} , then calculated GPP as the difference between bias-corrected R_{eco} and NEE. In summary, FluxPulse re-estimates R_{eco} during pulse events by forcing R_{eco} on the first day of pulse events to approximate maximum NEE on the same day, then calculating R_{eco} in the following days using the decay rate k , which is computed empirically for each pulse event (equation (4)). A correction factor β , which was written as a decay function and calculated on a daily scale, was applied to all half-hourly night-time method R_{eco} ($\text{Reco}_{\text{NTmethod}}$) from the same day to obtain the half-hourly FluxPulse R_{eco} ($\text{Reco}_{\text{FluxPulse}}$) (equation (3))

$$\text{Reco}_{\text{FluxPulse},i,j} = \beta_i \times \text{Reco}_{\text{NTmethod},i,j}, \quad (3)$$

where $\text{Reco}_{\text{FluxPulse},i,j}$ is the FluxPulse bias-corrected R_{eco} on day i and half-hour j , and β_i is the correction factor applied to all $\text{Reco}_{\text{NTmethod},i,j}$ on day i and half-hour j , in which $\text{Reco}_{\text{NTmethod},i,j}$ is R_{eco} obtained from the NT method. β decays over time to represent the decaying characteristics of R_{eco} during pulse events^{13,15} (equation (4) and Fig. 1) and is calculated as

$$\beta_i = \frac{\text{NEE}_{98^{\text{th}}}}{\text{Reco}_{\text{NTmethod},98^{\text{th}}}} \times e^{-k \times (i-1)}, \quad (4)$$

where $\text{NEE}_{98^{\text{th}}}$ and $\text{Reco}_{\text{NTmethod},98^{\text{th}}}$ are, respectively, the 98th percentile of half-hourly NEE and night-time method R_{eco} on day 1 of the pulse event (equation (4)). The 98th percentile values were chosen to select the maximum NEE and night-time method R_{eco} during pulse events. On day 1 ($i = 1$), β_i is equal to the ratio between the 98th percentile NEE and the 98th percentile night-time method R_{eco} (equation (4)), hence for $\text{Reco}_{\text{NTmethod},i,j}$ equals to $\text{Reco}_{\text{NTmethod},98^{\text{th}}}$, $\text{Reco}_{\text{FluxPulse},i,j}$ is equal to the 98th percentile NEE. This bias-correction method reduces the bias of the NT-method R_{eco} and forces the maximum of FluxPulse R_{eco} to approximate maximum NEE during pulse events, given R_{eco} should be theoretically larger or equal to NEE (equation (3)).

In equation (4), k is theoretically the decay rate of each pulse event. Since the half-hourly data are often noisy, and we do not have direct, continuous measurements of R_{eco} or R_h , the decay rate k was estimated empirically as follows (Extended Data Figs. 5 and 6). First, we fitted a first-order kinetic equation similar to equation (1) to the array $\text{NEE}_{68^{\text{th}}}$, $\text{NEE}_{68^{\text{th}}}$, $\text{NEE}_{68^{\text{th}}}$... $\text{NEE}_{68^{\text{th}}}$ in which $\text{NEE}_{p^{\text{th}}}$ is the p th percentile of all half-hourly NEE on day i . The 68th percentile was chosen to reflect values within ± 1 standard deviation from the mean on each day. If the nonlinear fit is statistically significant ($P < 0.1$), then the estimate of k is accepted. If the nonlinear fit is not statistically significant, we fitted a first-order kinetic equation to an alternative array: NEE_1 , NEE_2 , NEE_3 ... NEE_i , in which NEE_i is the mean of all half-hourly NEE in day i from all pulse events in a specific site. We validated FluxPulse R_{eco} against NEE at night, which was assumed to approximate R_{eco} due to the lack of photosynthesis at night (Extended Data Figs. 5 and 6). We only selected pulse events that lasted longer than 3 days for validation to minimize the effect of random error and the potential introduction of uncertainties due to low-turbulence conditions.

It is worth noting that in equation (4), by using NEE to set FluxPulse R_{eco} during pulse events, we implicitly assume that GPP during the first

day of rainfall is minimal due to the prepulse dry periods. Even though this is not the case in most sites (Extended Data Fig. 7), the assumption is necessary as there are no direct measurements of GPP and the partitioning methods available cannot provide reliable GPP estimates during pulse events. Importantly, this assumption is conservative and does not inflate the magnitude or importance of the main results. If GPP is positive on the first day of the pulse event, the bias in R_{eco} is expected to be even larger, as $R_{\text{eco}} = \text{NEE} + \text{GPP}$. Our reported bias values therefore probably underestimate the true bias associated with the partitioning methods assessed during pulse events (Fig. 2a,b).

Bias-correct carbon fluxes during prepulse periods. In addition to being biased during pulse events, the NT partitioning method is biased in the days leading up to pulse events (Fig. 2c and Extended Data Fig. 5). This is due to the fact that the NT method uses a moving window to estimate R_{ref} (equation (2)), and this window overlaps with the pulse event periods in the days before the pulse event. We therefore bias-corrected R_{eco} and GPP before the pulse events to reduce the prepulse biases (Fig. 2c, Extended Data Fig. 5 and Methods). We selected the prepulse period as n days before the pulse events (n being the event length) and used a 2-day window to estimate R_{ref} instead of a 4-day window used in the night-time method³⁴ (equation (2)). Only prepulse periods that do not overlap with previous pulse events were selected for correction.

FluxPulse technical assessment. Validation against the NEE measurements at night is complicated by the fact that rain events greatly reduce data quality for eddy-covariance measurements, particularly for open-path sensors. The three sites in the database use enclosed CO_2 sensors (ES-LM1, ES-LM2 and ES-Abr), however, which are robust under rainfall conditions⁷¹. Evaluation against night-time observations for those three sites shows that FluxPulse reduces the median R_{eco} bias from 27.0% to 0.7% (Extended Data Fig. 5). Across 82% of all studied sites, FluxPulse reduces biases significantly from the night-time method ($P < 0.001$, one-tailed paired t -test) and successfully captures the pulse temporal dynamics (Extended Data Figs. 5 and 6).

Hydrologic characteristics

Carbon flux characteristics during pulse events are strongly affected by the level of soil dryness and rainfall rewetting intensities^{14,15}. We tested two different indices for rewetting intensity: the EF and precipitation (P) (Extended Data Fig. 4). The rewetting intensity (ΔEF or ΔP) is the difference in either EF or P between 2 days after and 2 days before the rainfall. We calculated the antecedent water availability, denoted as prepulse EF, as the mean of EF 15 days before rainfall. A lower prepulse EF indicates a drier antecedent period. We took the average EF in 15 days before the pulse events happened but not a longer or shorter period, because a longer time window could include a mix of growing season and non-growing season, and a shorter period does not fully represent the history of water stress.

Soil characteristics

Microbial response to rewetting is influenced not only by water dynamics but also by soil properties^{15,72}. We extracted soil chemical and physical properties such as soil pH, texture and SOC from SoilGrids (www.soilgrids.org), which is a 250-m resolution interpolated global map of soil properties⁷³.

Analysis of spatiotemporal drivers of rain-induced carbon losses

We used the Lindeman, Merenda and Gold (LMG) method in the R package ‘relaimpo’ to assess the importance of spatial drivers such as soil characteristics, vegetation uptake, and climatic conditions on the intensity and decay rate of NEE during pulse events (Fig. 3). The LMG method is robust to multicollinearity between variables and provides confidence interval estimation via bootstrapping⁷⁴. Regarding the

temporal drivers of rain-induced carbon losses, we used the feature importance scores obtained from using random forests to label pulse events (‘Identification of pulse events’ section).

To analyse the directional relationship between site-specific pulse intensity/decay rate and each predictor, we used partial dependence plot, which effectively considers confounding effects between multiple predictors and the target variable, without assuming a linear relationship such as the Pearson correlation (Extended Data Figs. 8 and 10).

Data availability

Eddy-covariance sites are obtained from FLUXNET (www.fluxnet.org), AmeriFlux (www.ameriflux.lbl.gov) and ICOS (Integrated Carbon Observation System) Warm Winter 2020 (www.icos-cp.eu). Soil properties are obtained from Soilgrids (www.soilgrids.org). The manual pulse labelling dataset, FluxPulse bias-corrected eddy-covariance datasets, and datasets used to generate the figures are available via GitHub at <https://github.com/ngocnguyen99/FluxPulse>. Source data are provided with this paper.

Code availability

Related codes are publicly available via GitHub at <https://github.com/ngocnguyen99/FluxPulse>. The FluxPulse visualization tool is available via Shiny Applications at https://ngoc-nguyen-ucberkeley.shinyapps.io/fluxpulse_visualization/.

References

- Pastorello, G. et al. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Sci. Data* **7**, 225 (2020).
- Warm Winter 2020 Team and ICOS Ecosystem Thematic Centre. Warm Winter 2020 ecosystem eddy covariance flux product for 73 stations in FLUXNET-Archive format—release 2022-1 (version 1.0). *ICOS Carbon Portal* (2022); <https://www.icos-cp.eu/data-products/2G60-ZHAK>
- Vuichard, N. & Papale, D. Filling the gaps in meteorological continuous data measured at FLUXNET sites with ERA-Interim reanalysis. *Earth Syst. Sci. Data* **7**, 157–171 (2015).
- Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A. & Hegewisch, K. C. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci. Data* **5**, 170191 (2018).
- Shuttleworth, J., Gurney, R. J., Hsu, A. Y. & Ormsby, J. P. FIFE: the variation in energy partition at surface flux sites. In *Proc. IAHS Third Int. Assembly* 67–74 (1989).
- Grinsztajn, L., Oyallon, E. & Varoquaux, G. Why do tree-based models still outperform deep learning on tabular data? In *36th Conference on Neural Information Processing Systems (NeurIPS 2022)* https://papers.nips.cc/paper_files/paper/2022/file/0378c7692da36807bdec87ab043cdadc-Paper-Datasets_and_Benchmarks.pdf (NeurIPS, 2022).
- Breiman, L., Friedman, J., Olshen, R. A. & Stone, C. J. *Classification and Regression Trees* (Chapman and Hall/CRC, 2017); <https://doi.org/10.1201/9781315139470>
- Lloyd, J. & Taylor, J. A. On the temperature dependence of soil respiration. *Funct. Ecol.* **8**, 315–323 (1994).
- Nguyen, N. FluxPulse v1.0: model rain-induced carbon pulses at an ecosystem scale. R Shiny https://ngoc-nguyen-ucberkeley.shinyapps.io/fluxpulse_visualization/ (2025).
- Burba, G. *Eddy Covariance Method for Scientific, Regulatory, and Commercial Applications* (LI-COR Biosciences, 2022).
- Singh, S. How the Birch effect differs in mechanisms and magnitudes due to soil texture. *Soil Biol. Biochem.* **179**, 108973 (2023).
- Poggio, L. et al. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *SOIL* **7**, 217–240 (2021).

74. Lindeman, R. H., Merenda, P. F. & Gold, R. Z. *Introduction to Bivariate and Multivariate Analysis* Vol. 4 (Scott, Foresman Glenview, 1980).

Acknowledgements

This research received support through Schmidt Sciences, LLC., with a research exchange experience to the European Commission, Joint Research Centre in Ispra, Italy supported by the NSF FLUXNET AccelNet programme (award no. 2113978). M.B. acknowledges support from the US Department of Agriculture National Institute of Food and Agriculture (award no. 2023-67012-40086). D.P. thanks the support of the ITINERIS project (ID IRO00032) funded by EU—Next Generation EU Mission 4—Investment 3.1. T.F.K. acknowledges support from a DOE Early Career Research Program (award no. DE-SC0021023) and NASA (award nos. 8ONSSC21K1705 and 8ONSSC20K1801). M.B. and T.F.K. acknowledge support from NASA (award no. 8ONSSC25K7327). M.R. and K.-H.C. acknowledge funding by the European Research Council (ERC) Synergy Grant Understanding and Modeling the Earth System with Machine Learning (USMILE) under the Horizon 2020 research and innovation programme (grant agreement no. 855187). We thank D. Miller from Cornell University for proof-reading the manuscript and providing the aridity index dataset based on the TerraClimate precipitation and potential evapotranspiration data. We also thank T.T.H. Nguyen, J. Verfaillie, R. Amundson and scientists at Max Planck Institute for Biogeochemistry, Germany for valuable feedback on the mechanisms of the soil carbon pulses. The authors thank the AmeriFlux, ICOS and FLUXNET communities for the unique data contributions.

Author contributions

N.B.N., T.F.K. and M.M. designed the study. N.B.N. carried out the analysis, prepared figures and drafted the manuscript. D.D.B., L.A.G.,

M.M., M.B., J.K.G., D.P. and T.F.K. advised on ecological concepts and methods. T.F.K., D.D.B., D.P. and M.B. advised on the bias-correction method. M.B., T.F.K. and M.M. advised on water availability indices and pulse characteristics. H.H.N., Q.M.N. and T.D.N. advised on the machine learning pulse detection algorithm. K.-H.C. and M.R. provided data of the DML and NN partitioning methods. All authors participated in discussions at various stages and contributed significantly to the interpretation of the results and reviewing/editing.

Competing interests

The authors declare no competing interests.

Additional information

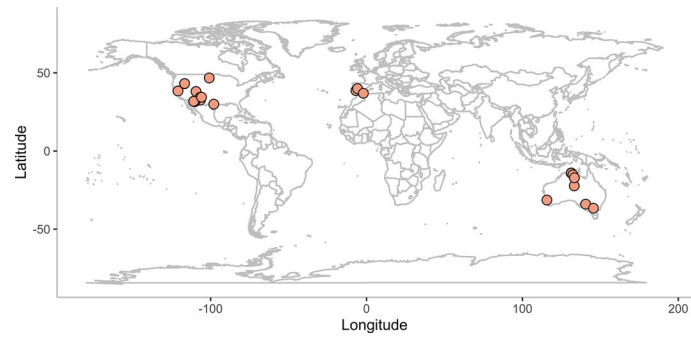
Extended data is available for this paper at <https://doi.org/10.1038/s41561-025-01754-9>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41561-025-01754-9>.

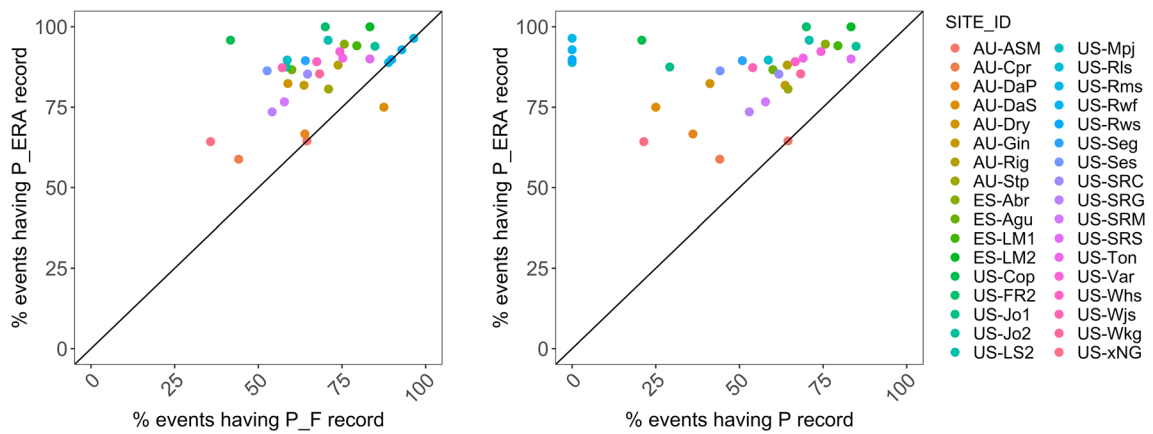
Correspondence and requests for materials should be addressed to Ngoc B. Nguyen or Trevor F. Keenan.

Peer review information *Nature Geoscience* thanks Andrew F. Feldman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Xujia Jiang and Tom Richardson, in collaboration with the *Nature Geoscience* team.

Reprints and permissions information is available at www.nature.com/reprints.

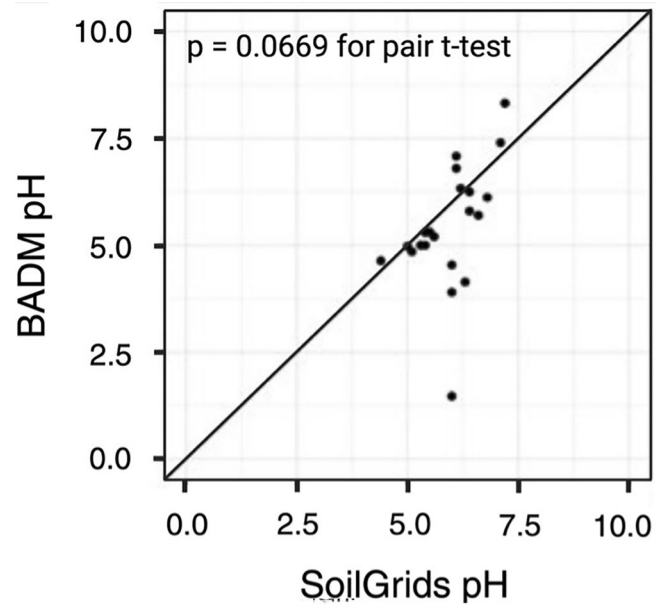


Extended Data Fig. 1 | Global map of sites included in this research. Thirty-four eddy-covariance sites were drawn as orange points.

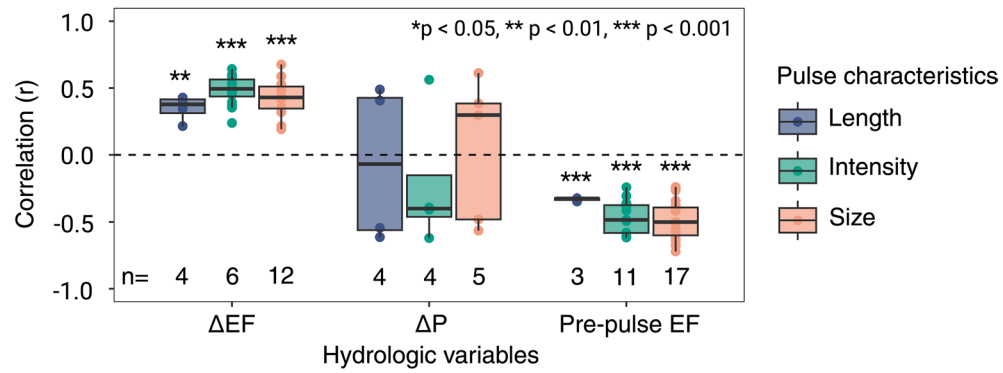


Extended Data Fig. 2 | Percent of pulse events that have precipitation data record in 34 sites. Percent of pulse events that have precipitation data record (P, P_F, and P_ERA columns in FLUXNET database) within ± 1 day from the start

day of the events in 34 sites (0.001 mm was set as the threshold for precipitation occurrence). P_ERA aligned most accurately with the occurrence of pulse events, whereas P_F and P records were absent for several of these events.

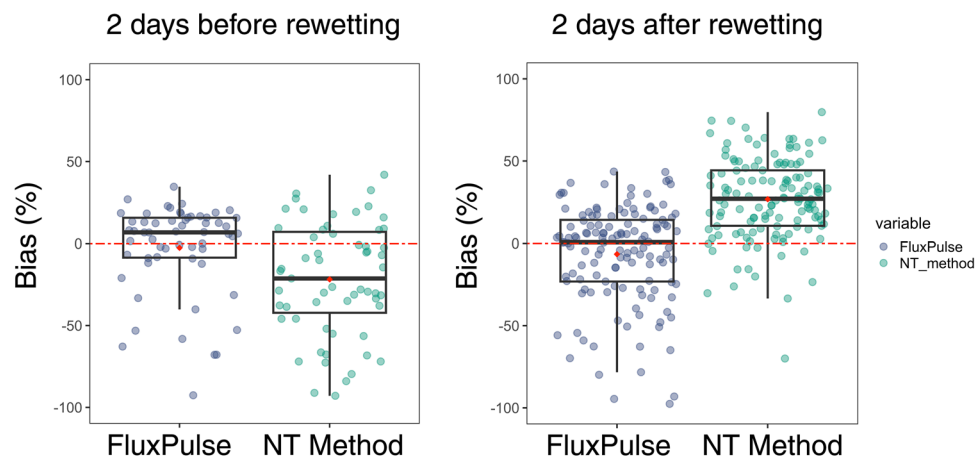


Extended Data Fig. 3 | Comparison of SoilGrids soil pH and in-situ measurements. Soil pH obtained from Soilgrids was compared to soil pH obtained from the BADM (Biological, Ancillary, Disturbance and Metadata) of AmeriFlux & ICOS sites.



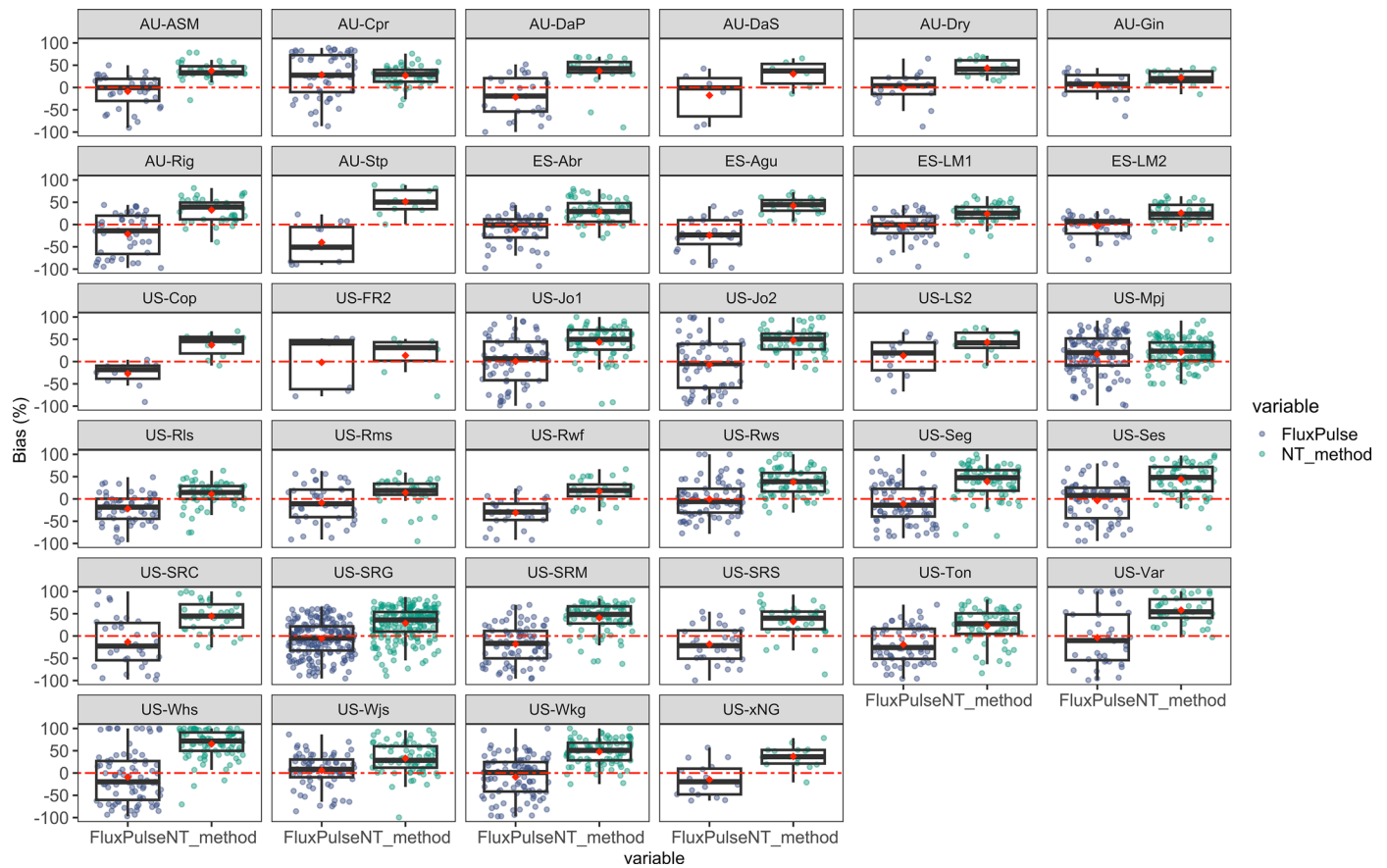
Extended Data Fig. 4 | Pearson correlation between pulse characteristics (Length, intensity, and size) and hydrologic variables (rewetting intensity (ΔEF , ΔP), antecedent water availability (Pre-pulse EF)) for each site. Only sites with a significant linear relationship with climatic drivers are shown as points

($p < 0.05$). Asterisks indicate the level of two-tailed t-test significance for whether the mean of the distribution is different than 0. The boxes represent interquartile range (IQR) marked with 25th, 50th, 75th percentile and whiskers extending to 1.5 times the IQR. n indicates the number of points in each boxplot.



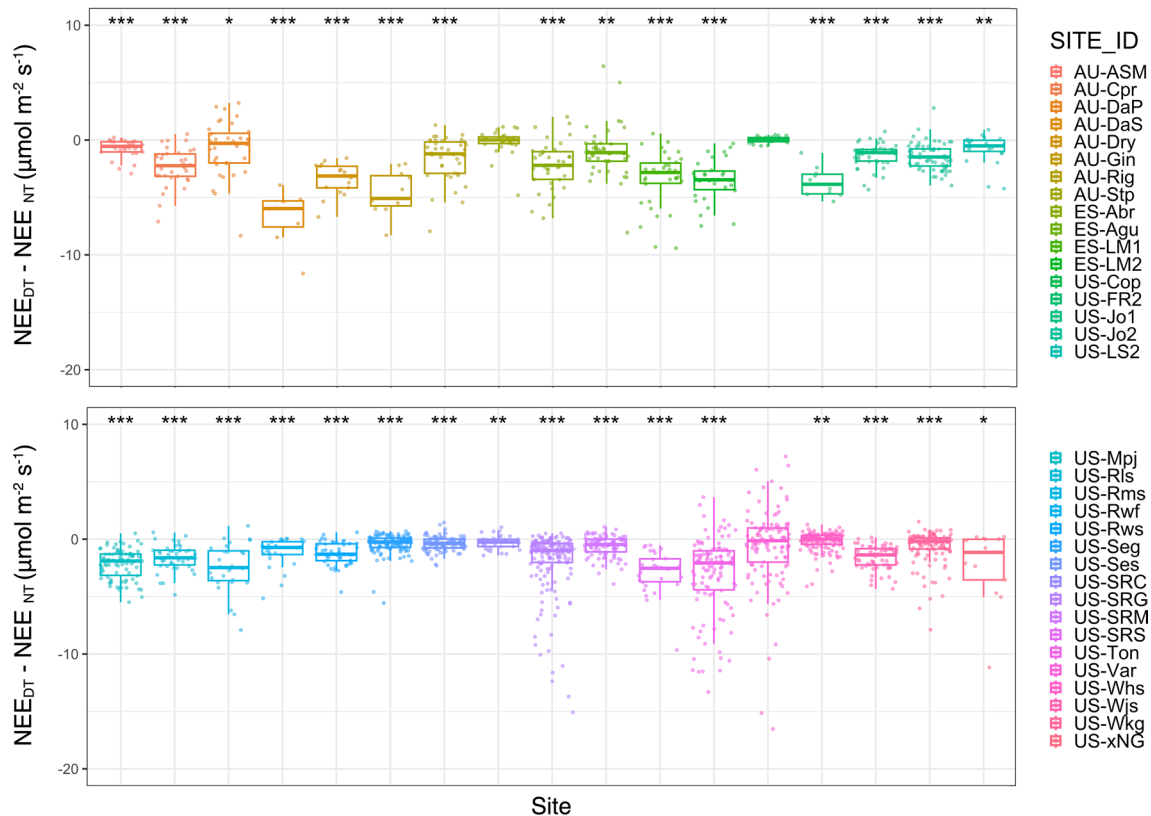
Extended Data Fig. 5 | Bias (%) of FluxPulse R_{eco} and Nighttime (NT)-method R_{eco} when comparing to nighttime NEE during two days before and after the start of pulse events at three sites that use an enclosed-path system (ES-LM1, ES-LM2, and ES-Abr). Bias (%) is calculated as $(x-y)/x \times 100\%$, in which x is the 68th percentile (1 standard deviation) of half-hourly Nighttime NEE , and y is the 68th percentile of half-hourly modelled R_{eco} (either FluxPulse or NT-method R_{eco}) during two days before and after the start of pulse events. Note that there are

fewer points before pulse events than after because we only bias-corrected the pre-pulse periods that are not overlapped with previous pulse events. The pre-pulse period covers n days before each pulse event, in which n is the length of that pulse event. The boxes represent interquartile range (IQR) marked with 25th, 50th, 75th percentile and whiskers extending to 1.5 times the IQR. The red diamonds represent the mean of the boxplots.



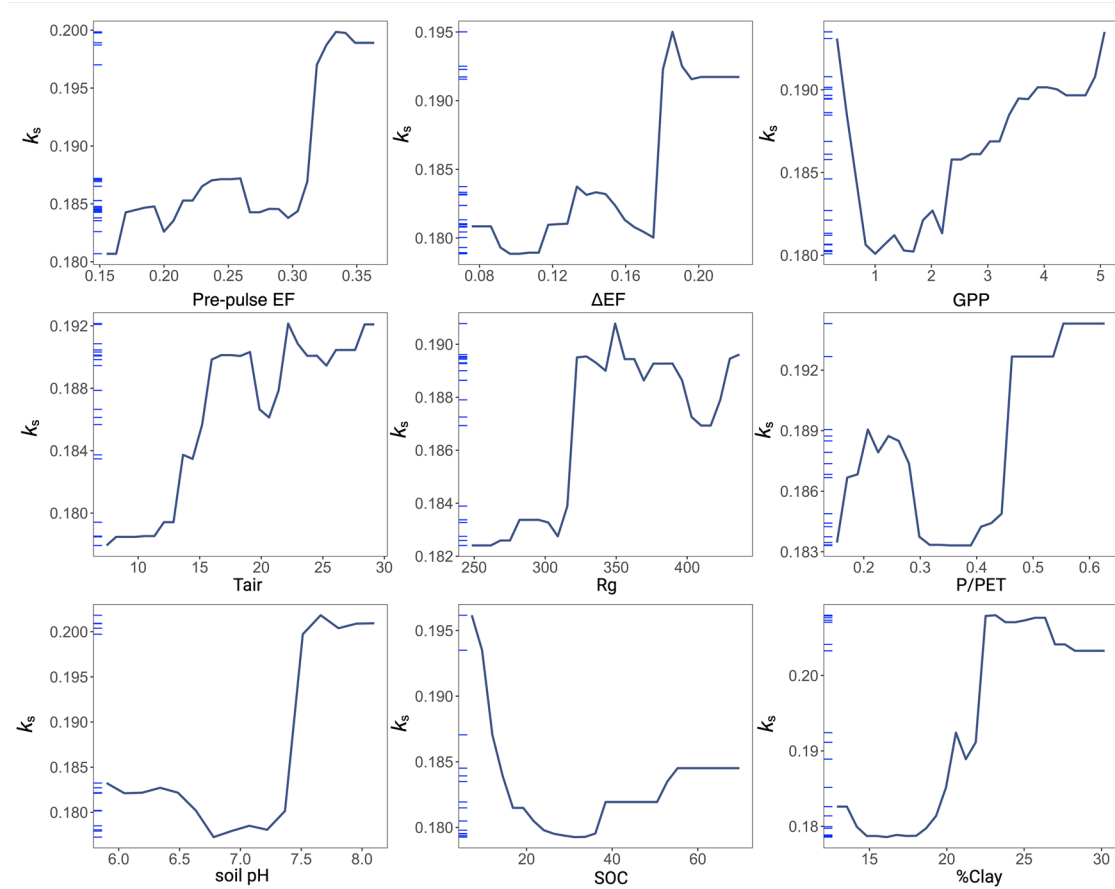
Extended Data Fig. 6 | Bias (%) of FluxPulse R_{eco} and Nighttime (NT)-method R_{eco} when comparing to nighttime NEE during two days after the start of pulse events at 34 sites. Bias (%) is calculated as $(x-y)/x \times 100$, in which x is the 68th percentile (1 standard deviation) of half-hourly Nighttime NEE, and y is the

68th percentile of half-hourly modelled R_{eco} (either FluxPulse or NT-method R_{eco}) during two days after the start of pulse events. The boxes represent interquartile range (IQR) marked with 25th, 50th, 75th percentile and whiskers extending to 1.5 times the IQR. The red diamonds represent the mean of the boxplots.



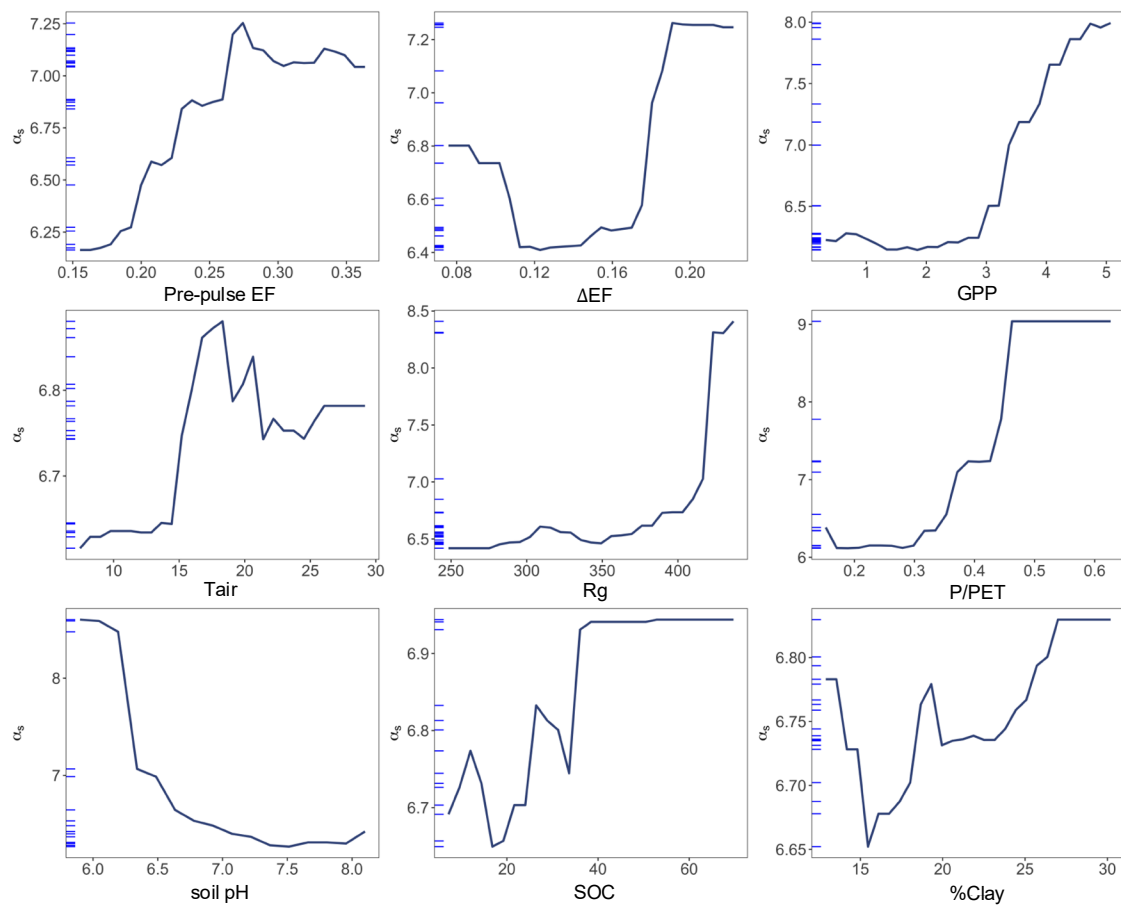
Extended Data Fig. 7 | Difference between half-hourly *NEE* during daytime and nighttime for the period of two days before each pulse event. Difference between half-hourly *NEE* during daytime and nighttime ($NEE_{DT} - NEE_{NT}$) for the period of two days before each pulse event. $NEE_{DT} - NEE_{NT}$ could indicate the presence of vegetation activity before pulse events, since the more negative is $NEE_{DT} - NEE_{NT}$, the more likely vegetation is active ($NEE_{DT} - NEE_{NT} = (R_{ecoDT} - GPP)$)

$-R_{ecoNT} = (R_{ecoDT} - R_{ecoNT}) - GPP$. Assuming that $R_{ecoDT} - R_{ecoNT}$ is positive since R_{eco} increases with temperature, then the observed negative $NEE_{DT} - NEE_{NT}$ must be due to positive *GPP*. Asterisks indicate the level of one-tailed t-test significance for whether the mean of the distribution is greater than 0 (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). The boxes represent interquartile range (IQR) marked with 25th, 50th, 75th percentile and whiskers extending to 1.5 times the IQR.



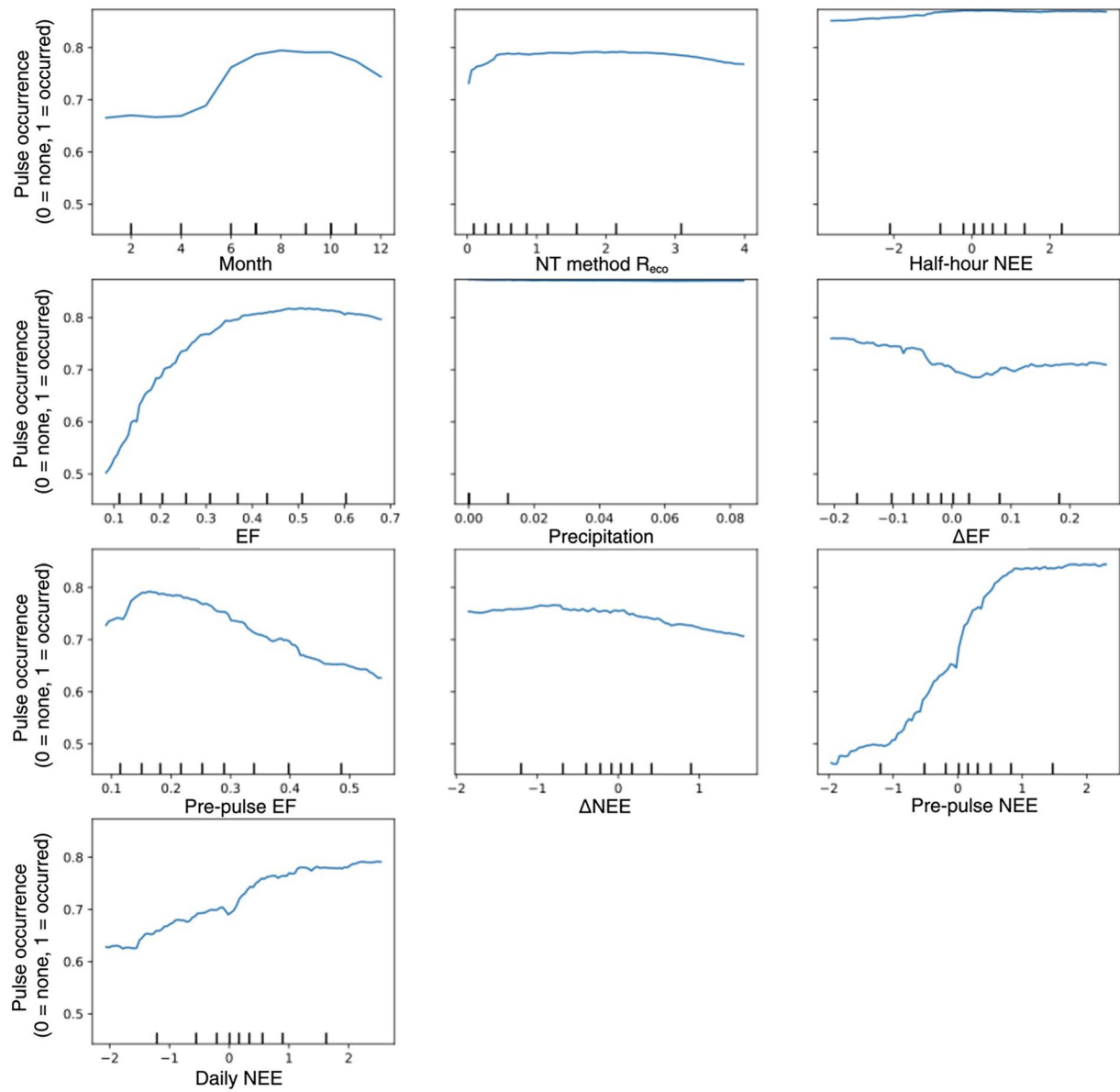
Extended Data Fig. 8 | Partial Dependence Plots for the responses of site-specific decay rate to nine predictors. Partial Dependence Plots for the responses of site-specific decay rate (k_s) to nine predictors. Predictors related to soil characteristics are clay percentage ($\%Clay$), soil organic carbon from 0–5 cm (SOC) ($g\ kg^{-1}$), and soil pH . The predictor representing vegetation productivity

is annual GPP obtained from the Nighttime (NT) partitioning method ($\mu mol\ CO_2\ m^{-2}\ s^{-1}$). Predictors representing hydrologic conditions are aridity index (P/PET), rewetting intensity (ΔEF), and antecedent water availability (Pre-pulse EF). Predictors representing climatic conditions are global shortwave radiation (R_g) ($W\ m^{-2}$), and air temperature (T_{air}) ($^{\circ}C$).



Extended Data Fig. 9 | Partial Dependence Plots for the responses of site-specific pulse intensity to nine predictors. Partial Dependence Plots for the responses of site-specific pulse intensity (α_s) to nine predictors. Predictors related to soil characteristics are clay percentage ($\%Clay$), soil organic carbon from 0–5 cm (SOC) ($g\ kg^{-1}$), and soil pH . The predictor representing vegetation

productivity is annual GPP obtained from the Nighttime (NT) partitioning method ($\mu mol\ CO_2\ m^{-2}\ s^{-1}$). Predictors representing hydrologic conditions are aridity index (P/PET), rewetting intensity (ΔEF), and antecedent water availability (Pre-pulse EF). Predictors representing climatic conditions are global shortwave radiation (R_g) ($W\ m^{-2}$), and air temperature (T_{air}) ($^{\circ}C$).



Extended Data Fig. 10 | Partial Dependence Plots for the responses of pulse occurrence to temporal predictors. Partial Dependence Plots for the responses of pulse occurrence (value of 1 indicates that a pulse event occurred, while 0 represents no likelihood of a pulse event) to temporal predictors (see Methods and Fig. 4c).