Observational partitioning of water and CO_2 fluxes at NEON sites: a five-year dataset of soil and plant components for spatial and temporal analysis

Einara Zahn[1] and Elie Bou-Zeid[1]

[1]Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, USA

Correspondence: Einara Zahn (einaraz@princeton.edu)

Abstract. Long-term time series of transpiration, evaporation, plant net photosynthesis, and soil respiration are essential for addressing numerous research questions related to ecosystem functioning. However, quantifying these fluxes is challenging due to the lack of reliable and direct measurement techniques, which has left gaps in the understanding of their temporal cycles and spatial variability. To help address this open challenge, we generated a dataset of these four components by implementing five (conventional and novel) approaches to partition total ET and CO_2 fluxes into plant and soil fluxes across 47 NEON sites. The final dataset (https://doi.org/10.5281/zenodo.12191876) spans a five-year period and covers various ecosystems, including forests, grasslands, and agricultural terrain. This is the first comprehensive dataset covering such a wide spatial and temporal distribution. Overall, we observed good agreement across most methods for ET components, increasing the confidence in these estimates. Partitioning of CO_2 components, on the other hand, was found to be less robust and more dependent on prior knowledge of water-use efficiency. This highlights some limitations of these present methods that we discuss, emphasizing the broader challenge posed by the lack of an accurate reference method to validate against. Despite these limitations, this dataset has several potential applications, especially in addressing critical questions regarding the response of ecosystems to extreme weather events, which are expected to become more severe and frequent with climate change.

1 Introduction

Plant transpiration and photosynthesis are important components of the global water and CO_2 cycles. At the ecosystem level, these quantities reflect the complex interactions within the soil-vegetation-atmosphere continuum. Given the complexity of these interactions and the challenges associated with measuring or modelling leaf-atmosphere exchanges, many open questions remain. For instance, the competing effects of rising CO_2 and temperature on plant water-use efficiency (Mengis et al., 2015; Kirschbaum and McMillan, 2018; Dusenge et al., 2019; Baslam et al., 2020; Wang et al., 2022) is still under debate, as well as their response and resilience under more frequent and intense droughts and changes in soil moisture conditions (Maxwell and Condon, 2016; Lesk et al., 2021). In addition to the challenges of investigating such non-linear processes, one of the main barriers to advancing research in this area has been the lack of long-term observations and/or reliable estimates of plant transpiration and CO_2 assimilation over various types of vegetation and soils.

Eddy-covariance (EC) is a reliable approach to continuously monitor evapotranspiration (ET) and net CO_2 (Fc) fluxes across ecosystems in a standardized manner. However, the separate measurement of the ecosystem individual components, namely plant transpiration (T) and soil evaporation (E), as well as plant gross primary production (GPP) and ecosystem respiration (R_{eco}), is still difficult. Traditional measurement techniques, such as sap-flow and lysimeters, often lack the temporal and spatial resolution necessary for capturing ET flux components comprehensively (Kool et al., 2014; Stoy et al., 2019). Other techniques, such as soil chambers (Kool et al., 2014), stable isotopes (Good et al., 2014), and carbonyl sulfide as a tracer for plant fluxes (Wohlfahrt et al., 2011), are not yet deployable at large spatial and temporal scales. Additionally, all of these methods carry their own uncertainties, can be logistically challenging to implement, and are often costly to scale up. Partitioning of CO_2 fluxes into GPP and R_{eco} can also be done based on night-time CO_2 flux extrapolation (Reichstein et al., 2005) and light-response curves (Lasslop et al., 2010); this method has been widely implemented and is part of FLUXNET products (Pastorello et al., 2020). Nonetheless, it cannot partition ET components. As a result, there is a scarcity of datasets representing temporal and spatial variability of both ET and CO_2 flux components across a broad range ecosystems and weather conditions.

An alternative approach to obtain long-term time series of flux components is through the partitioning of evapotranspiration and net CO_2 exchanges into their respective plant and soil components based strictly on the analysis of their turbulent perturbations signals. This approach is the focus of this paper, which applies five recently developed and tested partitioning methods to disentangle the contribution of transpiration, evaporation, net photosynthesis and soil respiration to the total ecosystem fluxes measured across all the EC towers of National Science Foundation's National Ecological Observatory Network (NEON). The methods employed in this study have been rigorously evaluated in previous studies that combined experimental data from EC sites and virtual EC data from large-eddy simulations. Although as for all partitioning approaches the validation is hindered by the lack of a universally-accepted accurate standard method, previous research has highlighted conditions under which these turbulence methods agree and are more reliable, as well as conditions when their partitioning estimates may have larger errors. In the dataset we produce in this work, we consolidate the results from five partitioning methods and calculate their ensemble average, providing a measure of uncertainty for each flux period at each site.

With a total of 47 sites spanning diverse ecosystems in the United States, this dataset can be an invaluable resource for future investigation of the spatial variability and ecological and meteoclimatic drivers shaping flux dynamics across different landscapes. It has the potential to contribute to the advancement of our understanding of ecosystem processes and their responses to current and future climate, in turn better informing ecosystem management and climate change mitigation and adaptation practices.

2 Study area and input data

The National Ecological Observatory Network (NEON) is a comprehensive initiative designed to collect long-term open-access environmental observations in the continental United States (Metzger et al., 2019). It operates a network of field sites strategically located to capture the diversity of ecosystems and environmental conditions present in the country. One of the main features of NEON is the standardized collection of data across sites, including biological, hydrological and meteorological observations. As part of its effort to understand land-atmosphere exchanges of energy, water and CO_2, NEON has a total of 47 eddy-covariance towers spread across 24 states (see Table 1), where most sites are located in deciduous and evergreen forests (Figure 1).

For the purpose of implementing the partitioning methods, we downloaded instantaneous measurements from the eddy-covariance towers from 2019 to 2023 for all 47 locations. All sites were equipped with the same gas analyzer (LICOR LI7200) and sonic anemometers (Campbell Scientific, CSAT-3 3-D) collecting data at a frequency of 20 Hz. These time series were processed and prepared to be used by the five partitioning methods, as described in the next section. In addition to raw-data, we also downloaded the processed eddy-covariance data (product level 4) for the same period (National Ecological Observatory Network (NEON), 2024a, b), which include evapotranspiration and net ecosystem CO_2 fluxes as processed by eddy4R (Metzger et al., 2017). Fluxes computed by eddy4R include all standard corrections, such as high-frequency spectral correction, some of which are only applied to the final turbulent fluxes, and not the instantaneous time series. Thus, we use the total turbulent fluxes as quantified by NEON to rescale all flux components that we computed, T, E, R, and P, ensuring that they are consistent with the corrected fluxes reported by the network.

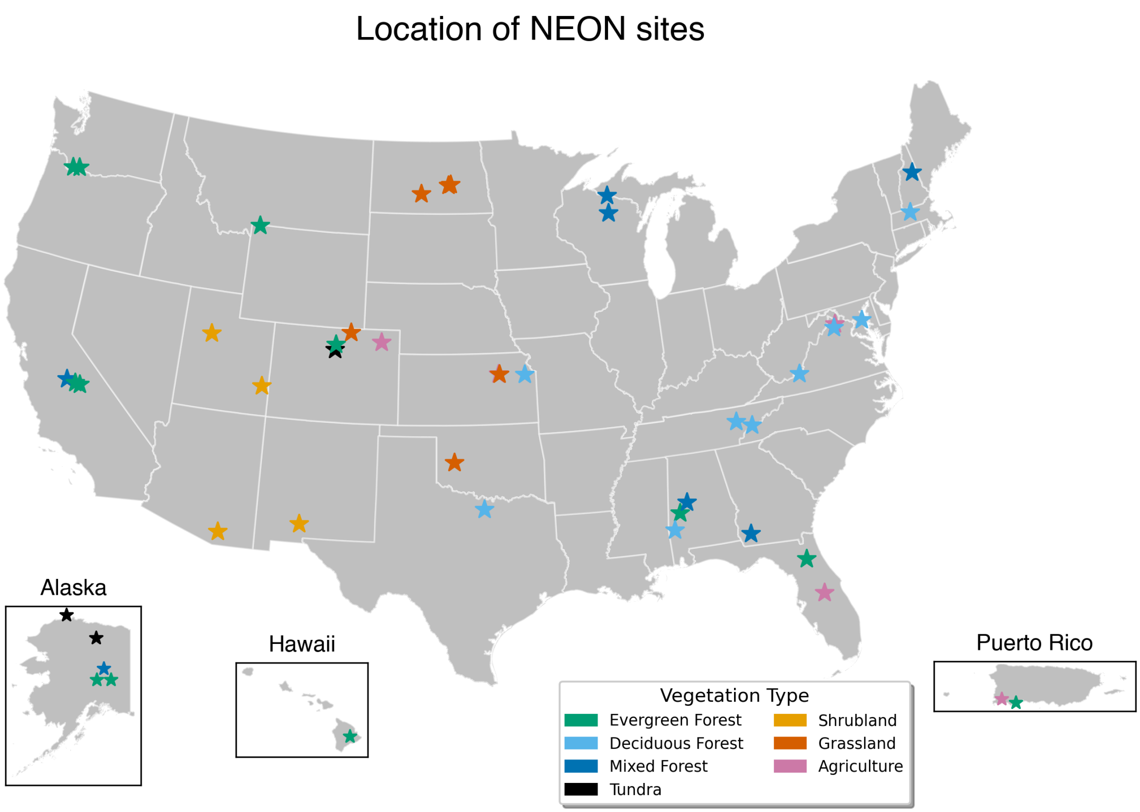

Figure 1. Location of the 47 eddy-covariance towers maintained by NEON. Note that Alaska, Hawaii, and Puerto Rico are not scaled proportionally to the continental US area.

As the last step of post-processing, a gap-filling algorithm was implemented to maximize the availability of flux partitioning estimates. Additional meteorological variables were downloaded using the R package neonUtilities (Lunch et al., 2024) and used as features to implement an Extreme Gradient Boosting (XGBoost) algorithm. The selected variables (inputs) included air temperature (National Ecological Observatory Network (NEON), 2024g, h), incoming solar radiation (National Ecological Observatory Network (NEON), 2024i, j), relative humidity (National Ecological Observatory Network (NEON), 2024e, f), photosynthetically active radiation (National Ecological Observatory Network (NEON), 2024c, d), and wind speed (National Ecological Observatory Network (NEON), 2024k, l). All quantities were measured at the top of the tower. The sensible heat flux and the friction velocity, included in the bundled data products, were also used. Due to the presence of long gaps and/or noisy measurements, precipitation and soil moisture were not included as features.

Table 1: Description of the 47 sites included in the present paper. In addition to the NEON site identification (ID) and state, the table also lists the geographical coordinates, the ratio between the eddy-covariance measurement height to mean canopy height, z/h, and the predominant vegetation type around the tower following Fiorella et al. (2021). The last two columns represent the availability of ET fluxes during daytime hours (when solar radiation $>$ 10 Wm^2) in the five years of record, and the availability of transpiration estimates, based on the partitioning approaches used here, in the same period. The range represents the availability according to the different methods (least available to most available). Availability computed after gap filling as described in section 3.4.2.

ID	**state**	**LAT**	**LON**	**z/h**	**Main vegetation type**	**ET availability (%)**	**T availability (%)**
ABBY	WA	45.76	−122.33	6.0	Evergreen Forest	61.3	58.7–61.2
BARR	AK	71.28	−156.62	30.0	Tundra Shrubland	33.7	24.7–33.7
BART	NH	44.06	−71.29	1.8	Mixed Forest	76.2	74.2–76.2
BLAN	VA	39.03	−78.04	2.0	Farm transitioning to forest	64.7	59.6–64.7
BONA	AK	65.15	−147.50	3.0	Mixed Forest	56.8	53.5–56.8
CLBJ	TX	33.40	−97.57	1.4	Deciduous Forest	74.3	72.6–74.3
CPER	CO	40.82	−104.75	30.0	Grassland	62.4	61.1–62.4
DCFS	ND	47.16	−99.11	8.0	Grassland	72.3	64.3–72.3
DEJU	AK	63.88	−145.75	3.6	Evergreen Forest	71.1	66.3–71.1
DELA	AL	32.54	−87.80	1.4	Evergreen Forest	54.1	48.9–54.0
DSNY	FL	28.13	−81.44	4.0	Pasture	69.2	65.7–69.2
GRSM	TN	35.69	−83.50	1.5	Deciduous Forest	65.7	62.9–65.7
GUAN	PR	17.97	−66.87	3.8	Evergreen Forest	50.7	48.2–50.7
HARV	MA	42.54	−72.17	1.7	Deciduous Forest	63.9	58.2–63.9
HEAL	AK	63.88	−149.21	30.0	Evergreen Forest	52.5	49.1–52.4
JERC	GA	31.19	−84.47	1.9	Mixed Forest	63.9	61.7–63.8

JORN	NM	32.59	−106.84	8.0	Desert Shrubland	76.9	73.3–76.9
KONA	KS	39.11	− 96.61	5.3	Cultivated Crops	76.7	70.9–76.7
KONZ	KS	39.10	− 96.56	8.0	Grassland	73.8	67.1–73.8
LAJA	PR	18.02	− 67.08	8.0	Pasture	46.8	45.6–46.7
LENO	AL	31.85	− 88.16	1.2	Deciduous Forest	43.5	37.1–43.4
MLBS	VA	37.38	− 80.52	1.5	Deciduous Forest	64.2	50.9–64.2
MOAB	UT	38.25	−109.39	40.0	Desert Shrubland	65.8	64.5–65.8
NIWO	CO	40.05	−105.58	13.3	Tundra/Alpine Shrubland	52.7	48.5–52.7
NOGP	ND	46.77	−100.92	20.0	Grassland	63.4	60.0–63.4
OAES	OK	35.41	− 99.06	5.3	Grassland	71.4	65.3–71.4
ONAQ	UT	40.18	−112.45	8.0	Sage Shrubland	71.4	67.6–71.3
ORNL	TN	35.96	− 84.28	1.4	Deciduous Forest	69.1	64.3–69.1
OSBS	FL	29.69	− 81.99	1.8	Evergreen Forest	55.5	51.3–55.5
PUUM	HI	19.55	−155.32	1.5	Evergreen Forest	57.5	41.3–57.5
RMNP	CO	40.28	−105.55	2.3	Evergreen Forest	78.7	74.4–78.7
SCBI	VA	38.89	− 78.14	1.7	Deciduous Forest	52.2	48.3–52.1
SERC	MD	38.89	− 76.56	1.7	Deciduous Forest	65.5	63.7–65.5
SJER	CA	37.11	−119.73	3.2	Mixed Forest	63.1	62.0–63.1
SOAP	CA	37.03	−119.26	2.1	Evergreen Forest	69.3	63.0–69.1
SRER	AZ	31.91	−110.84	4.0	Desert Shrubland	73.7	72.6–73.6
STEI	WI	45.51	− 89.59	2.3	Deciduous Forest	70.9	68.4–70.9
STER	CO	40.46	−103.03	2.7	Cultivated Crops	66.1	59.4–66.1
TALL	AL	32.95	− 87.39	1.4	Mixed Forest	63.0	58.5–63.0
TEAK	CA	37.01	−119.01	2.2	Evergreen Forest	63.6	61.4–63.5
TOOL	AK	68.66	−149.37	45.0	Tundra Shrubland	45.3	35.5–45.2
TREE	WI	45.49	− 89.59	1.5	Mixed Forest	76.1	69.8–76.1
UKFS	KS	39.04	− 95.19	1.9	Deciduous Forest	73.0	67.9–73.0
UNDE	MI	46.23	− 89.54	1.7	Mixed Forest	74.8	72.5–74.8
WOOD	ND	47.13	− 99.24	16.0	Grassland	68.9	64.8–68.9
WREF	WA	45.82	−121.95	1.4	Evergreen Forest	66.0	63.5–65.9
YELL	WY	44.95	−110.54	1.1	Evergreen Forest	57.4	53.2–57.4

3 Methods

Below, we describe the methods used to partition the data and the processing steps followed to generate the final dataset. A diagram summarizing all pre- and post-processing steps is shown in Figure 2. We implemented five methods to partition evapotranspiration into soil evaporation and plant transpiration, and net CO_2 flux into soil respiration and net plant photosynthesis. It is important to note that these quantities differ from GPP and ecosystem respiration (see Wohlfahrt and Gu (2015) for a detailed discussion on the different definitions of CO_2 flux components). Since these methods are based on principles of turbulent transport, they do not provide the information needed to separate soil respiration from plant respiration. Instead, the framework effectively partitions below-canopy fluxes, which include respiration and evaporation from soil, roots, and possibly the lower parts of tree trunks, and above-canopy fluxes, which encompass plant transpiration and net photosynthesis. Thus, none of these approaches allows for the distinction between above-ground respiration. Another limitation of these models is that they cannot distinguish between transpiration and evaporation of intercepted water on the leaves. Nonetheless, since EC data is often lost during and immediately following rainfall periods caused by poor data quality, many such periods are likely removed from the time series already. With these definitions in mind, we refer to net photosynthesis simply as photosynthesis (P) and plant ET simply at (T) in the following sections, while evaporation (E) and respiration (R) are considered to originate solely from the ground/soil level. We advise future users to carefully note the specific definitions of each component, as this is essential for making meaningful comparisons across different partitioning products.

3.1 Partitioning approaches

The partitioning methods implemented in this study are based on the idea of transport similarity between CO_2 and water vapor. It assumes that turbulent structures, or "eddies", simultaneously transport evaporation and respiration from the soil towards the sensor, where the respective concentrations of these "event samples" are measured. Similarly, turbulence also transports air parcels from the canopy, where net photosynthesis and plant transpiration take place. By separating the signature of soil versus canopy events in the time series, these methods try to infer the respective magnitude of each flux component. Note that while both evapotranspiration components are source of water vapor (i.e., they are positive fluxes), CO_2 fluxes are a combination of a source (soil respiration) and a sink (photosynthesis). This feature of the CO_2 flux is what creates this framework, since it works as a tracer that identifies the origin of water vapor. However, another consequence of this feature is that, while the components E and T are bounded by $E+T$, the magnitudes of $R>0$ and $P<0$ are not constrained by $F_c = R+P$. As a result, partitioning results for CO_2 components tend to be more uncertain. Nonetheless, as will be explained in the next section, two of the partitioning methods here implemented require the water-use efficiency, $W = P/T$, as an input. While the need for a parameterization for W is a disadvantage for the method, the inclusion of W adds a bound to P and R.

In total, five partitioning methods were implemented: the Flux-Variance Similarity (FVS) (Scanlon and Sahu, 2008; Scanlon and Kustas, 2010; Scanlon et al., 2019), the Modified Relaxed Eddy Accumulation (MREA) (Thomas et al., 2008), the Conditional Eddy Covariance (CEC) (Zahn et al., 2022), a modified CEC that also utilizes the water use efficiency (CECw) (Zahn et al., 2024), and the Conditional Eddy Accumulation (CEA) (Zahn et al., 2024). The advantage of these models is that they rely mostly on high-frequency eddy-covariance observations, requiring none or few input parameters. While MREA, CEC, and CEA are based solely on statistics computed from the high-frequency data, FVS and CECw additionally require the water-use efficiency as an input.

Three of the methods, FVS, MREA and CEC, have been extensively investigated in previous studies (Thomas et al., 2008; Klosterhalfen et al., 2019; Scanlon et al., 2019; Zahn et al., 2022). More recently, these three methods were tested using large-eddy simulations by Zahn et al. (2024), who also formulated and tested the CECw and CEA approaches, to probe the assumptions and the performance of all methods. Their simulation-based analyses indicated that all methods result in reliable results for a range of flux combinations that are expected in real ecosystems. While that study was not able to pinpoint conditions under which the accuracy of the results in field experiments can be guaranteed, it delineated flux combinations that may result in higher uncertainties, it quantified the impact of errors in W on the results of FVS and CECw, and it identified the most consequential assumptions in the various methods. The broad recommendations of that study are as follows:

1. The measurement height z, i.e., the height where the EC system is placed, should be as close as possible to the mean canopy height, h, to better distinguish the soil and plant signals. However, while $z/h < 3$ is a good recommendation, we note that the performance also depends on other factors such as canopy density and the specific method, as some approaches were found to perform well even outside this region.

2. All flux components from the soil (evaporation and respiration) and from the vegetation (transpiration and photosynthesis) should be non-negligible. This criterion arises from the water vapor-CO_2 coupling, which needs the CO_2 signal to determine the origin of the water vapor.

3. The (anti-)correlation between CO_2 and water vapor, $\rho_{c,q}$, should not be perfect, i.e., $|\rho_{c,q}| < 1$. Perfect correlation or anti-correlation indicate that turbulence fully mixed the scalars from the soil and from the canopy. In this case, the framework of all five partitioning models is no longer valid. Most often, high correlation is seen when measurements are too far from the canopy top, or when one of the flux components is negligible. Thus, this condition should be satisfied if points 1 and 2 are first addressed.

Overall, by comparing the outputs of the five models and identifying periods or sites where they consistently agree, we can increase our confidence in the results obtained, even without a "true" value for comparison. In addition, the range of predictions from the various methods can be used to quantify the uncertainty in the overall partitioning.

3.2 Data pre-processing

We downloaded and processed high-frequency eddy covariance turbulent exchange (ECTE) data using our in-house python routines. The following variables were extracted from the original datasets for each site: the mixing ratio of CO_2 and water vapor (r_c and r_q); the three components of the velocity field in the x, y, and z direction (u, v, and w); the sonic temperature (T_s); and the atmospheric pressure (P). In addition, sensor flags describing the quality of the measurements and signal were used for quality control (see Table 2 for a description of the used variables). Our algorithm was implemented following the pre-processing of daily files as follows:

1. Instantaneous measurements that were assigned 1 (poor quality) by any of the flags described in Table 2 were discarded;
2. CO_2 and H_2O mixing ratio observations were discarded if the signal strength indicator was smaller than 0.7;
3. Air density (kg m^{-3}), as well as mass concentrations of CO_2 and H_2O in mg m^{-3} and g m^{-3}, respectively, were computed;
4. Daily files with less than 50% of raw high-frequency data were discarded, with very few cases falling into this category. Files that passed this test were split into half-hour blocks of 36000 points each (at 20 Hz measurements);
5. Instantaneous observations were checked for physical plausibility, such as positive scalar concentrations;
6. Outliers were removed using a de-spike algorithm (Zahn et al., 2016);
7. The velocity field was rotated using double coordinate rotation;
8. CO_2 and H_2O mass concentrations, c and q, were lag-corrected, since the LI7200 is a close path sensor with tubing, using cross-correlation technique (Rebmann et al., 2012);
9. Turbulent fluctuations were computed for all quantities by removing the linear trend for all respective variables.

The outcome of these steps are 30-minute time series comprising cleaned turbulence data. In the next section we describe how these data were used to partition ET and F_c fluxes following the five methods described in section 3.1.

3.3 Implementation of partitioning methods

Since all partitioning methods are based on turbulence transport and similarity, their implementation require the computation of turbulent statistics, such as covariances and correlation coefficients (details of their implementation can be found in Zahn et al. (2022, 2024)). These statistics are computed for each 30-minute block. Using these turbulence statistics, we first implement the three methods that do not require extra inputs, namely CEC, CEA and MREA. For each partitioning method, we thus obtain T_{part}, E_{part}, P_{part}, and R_{part}, where the subscript "part" represents any of the three methods. For instance, for CEC we have T_{CEC}, E_{CEC}, P_{CEC}, and R_{CEC}.

The implementation of FVS and CECw additionally requires an estimate of the water-use efficiency, W. Note that in this context, W is defined as the ratio between net photosynthesis and transpiration, which again differs from other definitions. This quantity was also estimated for each 30-minute block by the following expression

$$W = 0.65\frac{\overline{c_c} - \overline{c_s}}{\overline{q_c} - \overline{q_s}}, \tag{1}$$

where $\overline{q_c}$ and $\overline{c_c}$ are H_2O and CO_2 atmospheric mean concentrations near the canopy, while $\overline{q_s}$ and $\overline{c_s}$ are the mean intercellular concentrations. The near-canopy concentrations ($\overline{q_c}$ and $\overline{c_c}$) are obtained using the logarithmic profile and mean concentrations from EC data, while q_s is calculated under the assumption of stomatal saturation (at air temperature) and a well-coupled leaf. $\overline{c_s}$, on the other hand, is the most challenging variable to obtain. Following previous studies (Wagle et al., 2021; Zahn et al.,

Table 2. Variables used for data pre-processing.

Variable	Name in NEON dataset	Description
r_c	rtioMoleDryCo2	mixing ratio $\mathrm{mol_{CO_2}/mol_{dryair}}$
r_q	rtioMoleDryH2o	mixing ratio $\mathrm{mol_{H_2O}/mol_{dryair}}$
u	veloXaxs	velocity in x direction (m/s)
v	veloYaxs	velocity in y direction (m/s)
w	veloZaxs	velocity in z direction (m/s)
T_s	tempSoni	sonic temperature (K)
P	presAtm	Atmospheric pressure (Pa)
-	ssiCo2	signal strength indicator for CO_2
-	ssiH2o	signal strength indicator for H_2O
-	qfSoniCode	Sensor error flag (Wrong embedded sensor code)
-	qfSoniComm	Sensor error flag (SDM communications error)
-	qfSoniData	Sensor error flag (No data available)
-	qfSoniSgnlHigh	Sensor signal flag (High signal amplitude)
-	qfSoniSgnlLow	Sensor signal flag (Low signal amplitude)
-	qfSoniSgnlPoor	Sensor signal flag (Poor signal lock)
-	qfSoniTemp	Sensor signal flag (Axes $\mathrm{T_{SONIC}}$ difference >4K)
-	qfSoniTrig	Sensor error flag (Sensor trigger source lost)
-	qfSoniUnrs	Sensor error flag (Sensor unresponsive)

2022), we implemented five models for $\overline{c_s}$, rendering a total of five estimates of W. These parameterizations are summarized in Table 3. For each 30-minute period, FVS and CECw were thus implemented five times, one for each parameterization of W. As a result, each flux component is estimated five times; for instance, following the nomenclature presented in Table 3, after implementing FVS we obtain $T_{\mathrm{FVS_{CC}}}$, $T_{\mathrm{FVS_{CR}}}$, $T_{\mathrm{FVS_{ID}}}$, $T_{\mathrm{FVS_{sD}}}$, and $T_{\mathrm{FVS_{OPT}}}$. These five estimates are then averaged to $T_{\mathrm{FVS_{MEAN}}}$ when at least one of the five outputs is available. The same averaging is done to obtain $E_{\mathrm{FVS_{MEAN}}}$, $P_{\mathrm{FVS_{MEAN}}}$, and $R_{\mathrm{FVS_{MEAN}}}$. Likewise, the same procedure is applied to CECw, which also takes W as an input. Our final dataset includes the outputs from all five models as well as the averaged values, as described in the next section.

3.4 Data post-processing

3.4.1 Data cleaning and flux rescaling

The last section described how all five partitioning methods were implemented to obtain estimates of soil and canopy fluxes at 30-minutes time interval. In post-processing, we first eliminated unrealistic high flux magnitudes, as well as counter-gradient fluxes, by discarding periods when any of the following conditions were identified:

Table 3. Description of all five parameterizations adopted for W. More details can be found in Wagle et al. (2021) and Zahn et al. (2022). All five parameterizations were used to partition fluxes by FVS and CECw, thus resulting in five partitioning results for each partitioning approach.

Method	Description	Variable name
Constant concentration	$\overline{c_s}$ is assumed a constant value	W_{CC}
Constant ratio	A constant ratio $k = \overline{c_s}/\overline{c_c}$ is assumed	W_{CR}
Linear	The ratio $k = \overline{c_s}/\overline{c_c}$ is a linear function of the vapor pressure deficit (VPD)	W_{lD}
Square root	The ratio $k = \overline{c_s}/\overline{c_c}$ is a square root function of VPD	W_{sD}
Optimum	W is calculated based on the optimization algorithm developed by Scanlon et al. (2019)	W_{OPT}

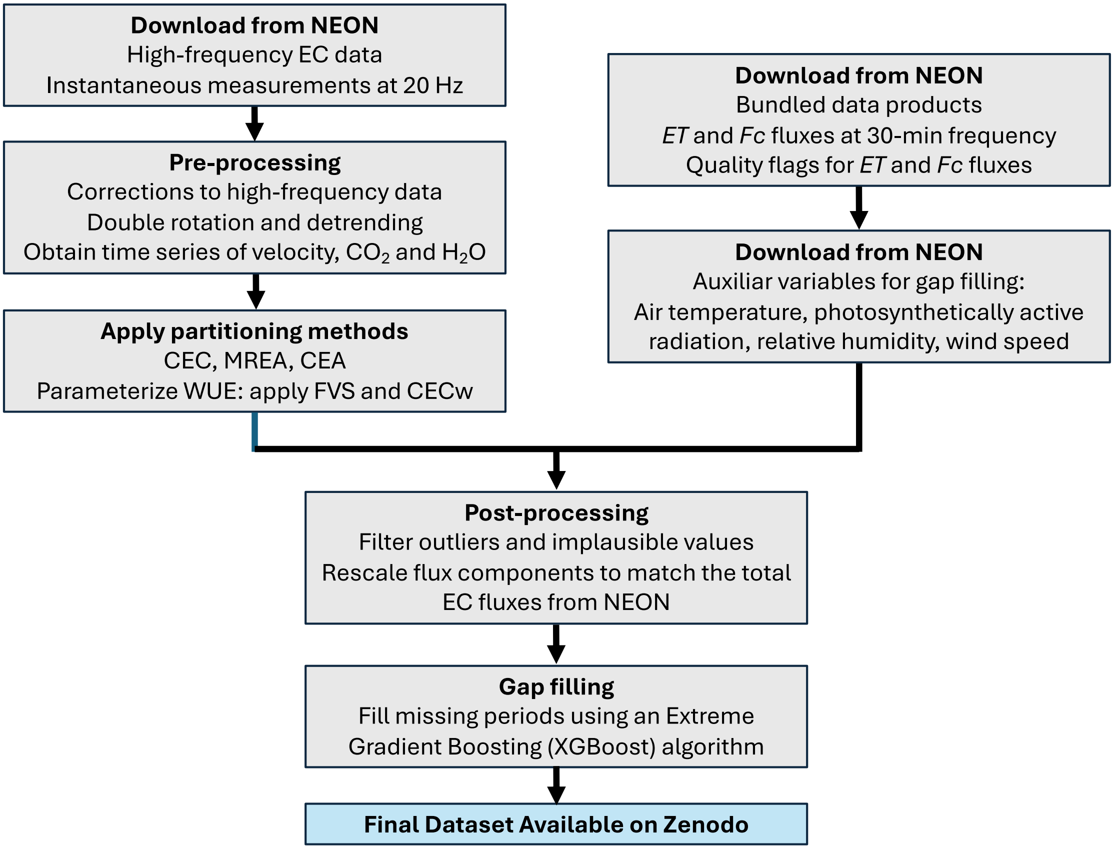

Figure 2. Diagram representing all steps of data processing used to generate the final dataset.

- $T_{\text{part}} < 0$ Wm^{-2} or $E_{\text{part}} < 0$ Wm^{-2};
- $P_{\text{part}} > 0$ mgm^{-2}s^{-1} or $R_{\text{part}} < 0$ mgm^{-2}s^{-1};
- $T_{\text{part}} + E_{\text{part}} > 1000$ Wm^{-2};
- $|P_{\text{part}} + R_{\text{part}}| > 10$ mgm^{-2}s^{-1};

As previously described, all partitioning methods relied on high-frequency time series for turbulent statistics and flux computation. Consequently, the resulting flux components lacked certain eddy-covariance corrections commonly applied to the final fluxes, including high-frequency spectral adjustments. To address this, we rescaled our fluxes based on those obtained from the NEON bundled data processed via eddy4R. We first computed correction factors for ET and F_c fluxes as

$$\text{cor}_{\text{ET}} = \frac{ET_{\text{NEON}}}{ET_{\text{part}}} \tag{2}$$

$$\text{cor}_{\text{Fc}} = \frac{Fc_{\text{NEON}}}{Fc_{\text{part}}}, \tag{3}$$

where ET_{NEON} and Fc_{NEON} are evapotranspiration and net CO_2 fluxes downloaded directly from NEON, while ET_{part} and Fc_{part} are the fluxes computed by our routines using the high-frequency data following the pre-processing described in section 3.2. Note that the total fluxes, $ET_{\text{part}} = T_{\text{part}} + E_{\text{part}}$ and $Fc_{\text{part}} = P_{\text{part}} + R_{\text{part}}$, are the same for all five partitioning methods. The corrections were then applied as

$$T_{\text{part,cor}} = \text{cor}_{\text{ET}} T_{\text{part}} \quad \text{and} \quad E_{\text{part,cor}} = \text{cor}_{\text{ET}} E_{\text{part}}, \tag{4}$$

$$P_{\text{part,cor}} = \text{cor}_{\text{Fc}} P_{\text{part}} \quad \text{and} \quad R_{\text{part,cor}} = \text{cor}_{\text{Fc}} R_{\text{part}}. \tag{5}$$

By rescaling our fluxes based on this comparison, we ensure that our fluxes are consistent with NEON products, mitigating potential discrepancies arising from variations in data processing methodologies. For simplicity, we drop the subscript ‘cor’ in the remainder of the text and in the data files.

Additionally, we utilized NEON’s bundled eddy-covariance products to assess the quality of both evapotranspiration and CO_2 flux data. Although our initial data preprocessing removed numerous periods with poor quality, we refrained from conducting further tests, such as assessing data stationarity, integral turbulence characteristics, and spectral analysis. Consequently, we integrated the flags as outlined in Metzger et al. (2022) into our final dataset. These flags are classified into two categories: 0, indicating compliance with all quality criteria, rendering the data valid and reliable for analysis, and 1, indicating failure to meet one or more quality tests, requiring cautious utilization or possible exclusion. Retaining the original flags in our dataset facilitates the identification of periods characterized by high-quality data. Nonetheless, we opted not to automatically discard data solely based on these flags, leaving such decisions to the discretion of subsequent users.

3.4.2 Gap filling

As a consequence of the inherent challenges associated with eddy-covariance measurements — such as sensor malfunction, data transmission, or even external factors such as rain and dust collection — flux time series are often incomplete. For our

Table 4. Success rate of each partitioning approach in finding a solution, expressed as a percentage of the available half-hour periods (i.e., only considering cases when the time series was available and the partitioning methods were implemented). The table shows the average and standard deviation across all sites before and after gap filling. For FVS and CECw, the convergence using each water-use efficiency parameterization is also shown.

Method	% periods with solution before gap filling (mean± std)	% periods with solution after gap filling (mean± std)
CEC	99.9 ± 0.1	100.0 ± 0.0
MREA	80.3 ± 7.6	91.4 ± 4.0
CEA	81.0 ± 8.6	97.8 ± 2.4
FVS-W_{CC}	51.9 ± 10.3	80.8 ± 21.6
FVS-W_{CR}	52.4 ± 10.3	80.0 ± 22.7
FVS-W_{lD}	48.8 ± 10.9	89.1 ± 6.8
FVS-W_{OPT}	50.9 ± 9.9	90.2 ± 10.8
FVS-W_{sD}	52.1 ± 10.4	86.5 ± 15.0
CECw-W_{CC}	86.2 ± 7.8	96.9 ± 3.6
CECw-W_{CR}	87.1 ± 7.5	96.9 ± 3.4
CECw-W_{lD}	72.0 ± 12.0	93.5 ± 5.2
CECw-W_{OPT}	49.5 ± 9.1	91.9 ± 5.8
CECw-W_{sD}	84.3 ± 8.0	96.1 ± 3.7

dataset, Table 1 shows the availability of ET fluxes during the five-year period for each site. Considering only daytime periods (selected based on incoming solar radiation greater than 10 Wm^2), it reveals that some sites have available fluxes only 34% of the time (BARR), while others have daytime ET estimates up to 77% of the five-year record (KONA).

Due to mathematical constraints and/or invalidity of physical assumptions, partitioning methods are not guaranteed to converge to a solution for every 30-minute time series, leading to further gaps in the record of flux components. As a consequence, the dataset generated as described in the previous sections contains missing periods, with the length of these gaps depending on the site and method. Table 4 illustrates that the methods with the lowest convergence rates—defined as the fraction of solutions found per available period—range from least to most solutions found as follows: FVS, CECw, MREA, CEA, and CEC. The lowest rate of convergence for the FVS method (on average 50%) is not surprising, and has been discussed in previous studies (Wagle et al., 2021; Zahn et al., 2022, 2024). Additionally, the low rate of solutions for FVS and CECw when combined with W_{OPT} is due to the model for this water-use efficiency itself not always converging (i.e., W_{OPT} was not available).

To increase the availability of periods with valid partitioning estimates for each method, we implemented Extreme Gradient Boosting (XGBoost) (Chen and Guestrin, 2016) for regression to predict transpiration. By training the algorithm to partition ET into E and T based on selected environmental features, the goal is to apply the trained algorithm to estimate T only when ET is available but a given partitioning method did not yield a valid estimate. That is, an XGBoost model is developed for each partitioning approach independently and then applied to fill the gaps in the time series of that approach. Additional machine learning algorithms, including neural networks and k-means clustering, were investigated but yielded similar or inferior results compared to XGBoost. Previous studies focusing on gap filling eddy-covariance data also observed superior performance by XGBoost (Huang and Hsieh, 2020; Irvin et al., 2021). These findings are corroborated by research indicating that tree-based models outperform deep learning for tabular data used in regression tasks (Grinsztajn et al., 2022). Thus, the selected algorithm is robust and a very suitable choice for the problem at hand.

Only daytime conditions were considered, a constraint imposed by selecting periods when the incoming solar radiation was > 10 Wm^2. In addition, we also ensured that only cases with positive ET were included. For regression purposes, we further filtered out periods where one or more input features were unavailable. As a result of these criteria, not every segment of the time series was suitable for gap filling. Nevertheless, this approach maximizes data availability of T and E, in particular during daytime, expanding the use of this dataset in future studies. Note that gap filling was not applied to F_c components. This decision was made due to the noisier nature of CO_2 fluxes. Furthermore, as discussed later, estimates for respiration and photosynthesis are less reliable across methods (larger variance between methods), and thus gap filling for these components was not deemed as robust as with ET components.

To maximize the periods with available data features, we only selected meteorological variables with small periods of missing data and high quality. Following tests on different meteorological variables, the following features were retained: ET, air temperature (T_{AIR}), photosynthetically active radiation (PAR), relative humidity (RH), vapor pressure deficit (VPD), mean wind speed (U), sensible heat flux (H), and friction velocity ($u_\star$). In addition, daily and seasonal cycles were represented as

$$X_{\text{diurnal}} = \frac{1}{2}\sin\left(\frac{2\pi}{24}\text{hour} - \frac{\pi}{2}\right) + \frac{1}{2}, \tag{6}$$

$$X_{\text{yearly}} = \frac{1}{2}\sin\left(\frac{2\pi}{365}\text{julian} - \frac{\pi}{2}\right) + \frac{1}{2}, \tag{7}$$

where X_{diurnal} and X_{yearly} represent the diurnal and yearly cycles corresponding to the hour of the day (in local time) and Julian day, respectively. While there are other variables that could be relevant for this problem, they were excluded due to poor data quality or availability. Nonetheless, using only the listed variables resulted in excellent performance metrics.

For each site and partitioning method — including the different combinations of FVS, CECw and water-use efficiency models taken individually— the dataset was divided into training (60%), validation (20%), and testing (20%) sets. Across different cases, the size of the training set varied between 1700 to 12000 points. To monitor the performance of the algorithm, we used the coefficient of determination, R^2, defined as

$$R^2 = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\overline{y})^2}, \tag{8}$$

where y are the observed values with average $\overline{y}$, while $\hat{y}$ are the predictions. This metric indicates how well the independent variables explain the variance in the dependent variable. Thus, the upper bound $R^2 = +1$ is expected only if the selected features explain 100% of the variance.

After experimenting with different hyperparameters, the same values for learning rate (0.01), maximum depth of a tree (5), and number of trees (1000) were used to train all sites and methods individually. Finally, gap filling was only performed for a particular case when its respective R^2 on the test set was greater than 0.7. Few cases had an R^2 below this threshold, which was found to correspond most often to datasets combining FVS and W_{CC} or W_{CR} over shrublands or tundra sites, combinations which resulted in the least amount of valid partitioning estimates (smallest training sets). Excluding these few exceptions, the majority of the cases featured $R^2 > 0.8$ (and bias < 1 Wm^{-2}) and were deemed reliable for gap filling. The final metrics for all models and sites can be found in the supplementary material. All auxiliary data used to implement the gap-filling algorithm is also available included in the final dataset.

As expected, the algorithm identified ET as the most relevant variable across all cases (approximately 80%). Note that in this context, we do not use ET to explain the variability in E or T in the sense of an environmental forcing. Instead, ET is used to teach the model how to partition the fluxes, effectively functioning as a machine-learning-based partitioning method. In section 6, we briefly explore feature importance for T and T/ET in the context of environmental forcings, where only environmental variables are included as features (excluding ET), and we investigate how much of these fluxes they can explain.

While the variability in feature importance was small across vegetation types, a larger variability was observed between methods. Notably, FVS combined with the water-use efficiency models W_{CC} and W_{CR} often attributed a lower importance to ET (as low as 27%) and higher importance to VPD (as high as 25%). Previous work (Wagle et al., 2023) has shown that these two water-use efficiency models are very sensitive to the choice of parameters, and greatly overestimate W compared to other models, particularly early in the morning. While investigating the possible causes of this discrepancy across water-use efficiency models is beyond the scope of this paper, deeper analyses will be enabled by this dataset.

Following the algorithm evaluation, transpiration was then estimated for missing data periods and defined as $T_{\mathrm{part}}^{\mathrm{xgb}}$, where the superscript 'xgb' refers to the the gap-filled variable using XGBoost. Only half-hour periods when $0 < T_{\mathrm{part}}^{\mathrm{xgb}} < ET$ were gap filled. Evaporation was then estimated as $E_{\mathrm{part}}^{\mathrm{xgb}} = ET - T_{\mathrm{part}}^{\mathrm{xgb}}$. For FVS and CECw, we also computed the average across their different water-use efficiency outputs, corresponding to $T_{\mathrm{FVS_{MEAN}}}^{\mathrm{xgb}}$ and $T_{\mathrm{CECw_{MEAN}}}^{\mathrm{xgb}}$.

The convergence rate to valid solutions after gap filling is shown in the last column of Table 4. Most notably, gap filling increased the presence of estimates from the FVS method for 80 to 90% of the periods when partitioning was applied. To further illustrate the completeness of the flux-partitioning record in the five-year dataset, Table 1 shows the availability of flux components as a fraction of the total number of half-hour periods in the record. Overall, all methods cover a similar temporal distribution of flux partitioning and are potential candidates for ensemble averaging.

4 Description of the final dataset

The final dataset is available for download at https://doi.org/10.5281/zenodo.12191876. It is organized in different folders for each site, with each site containing a CSV file for each method. This format is selected to be user-friendly and accessible in various programming languages and software. For FVS and CECw, in addition to their ensemble averages for different water-use efficiency parameterizations, we also include each individual output. This allows future users to compare the impact of water-use efficiency on the partitioning output.

All results are in UTC time, where it represents the start time of measurements for a half-hour period (i.e., starting at minute '00' or '30'). Each file contains all fluxes in continuous timestamps of 30 minutes. In cases where a specific variable was unavailable, the corresponding cell in the CSV file was left blank. A python script showing how to read the final data files is included with the dataset.

In each file, the following variables are included

- `Tpart`: transpiration [W m^{-2}]
- `Tpart_xgb`: transpiration after gap filling [W m^{-2}]
- `Ppart`: plant net photosynthesis [$mg_{CO_2} m^{-2} s^{-1}$]
- `Epart`: below canopy evaporation [W m^{-2}]
- `Epart_xgb`: below canopy evaporation after gap filling [W m^{-2}]
- `Rpart`: below canopy respiration [$mg_{CO_2} m^{-2} s^{-1}$]
- `Fc`: net CO_2 flux [$mgCO_2/m^2/s$]
- `ET`: evapotranspiration [W/m^2]
- `qfFinalCO2`: NEON final quality flag for CO_2 flux (0: `Fc` passed all quality assurance tests, 1: `Fc` flagged by one or more test)
- `qfFinalH2O`: NEON final quality flag for `ET` flux (0: `ET` passed all QA tests, 1: `ET` flagged by one or more quality assurance tests)

For flux components, "part" represents any of the methods. In addition, a separate file, **WUEinput.csv**, contains the water-use efficiency values from all five parameterizations, defined as follows

- `const_ppm` [$kg_{CO_2} (kgH_2O)^{-1}$]
- `const_ratio` [$kg_{CO_2} (kgH_2O)^{-1}$]
- `linear` [$kg_{CO_2} (kgH_2O)^{-1}$]
- `sqrt` [$kg_{CO_2} (kgH_2O)^{-1}$]
- `opt` [$kg_{CO_2} (kgH_2O)^{-1}$]

Finally, a file named **MeteorologicalVariables.csv** is also included, which contains all the auxiliary meteorological variables used for gap filling and feature importance analysis. This file is synchronized with the partitioning data file to contain the same records. The following variables are available

– `SoilMoisture`: soil moisture measured at the second level below the surface (depth varies by site) [$\mathrm{m^3 m^{-3}}$]

– `AirTemperature`: air temperature [°C]

– `SolarRadIn`: incoming solar radiation measured at the top of the tower [$\mathrm{Wm^{-2}}$]

– `RH`: relative humidity in %

– `Rain`: accumulated precipitation at the top of the tower during that 30 min period [mm]

– `PAR`: photosynthetically active radiation measured at the top of the tower [μmolesm^{-2}s^{-1}]

– `fluxTemp`: sensible heat flux at the top of the tower [$\mathrm{Wm^{-2}}$]

– `ustar`: friction velocity at the top of the tower [$\mathrm{ms^{-1}}$]

– `WindSpeed`: mean wind speed at the top of the tower [$\mathrm{ms^{-1}}$]

– `vpd`: vapor pressure deficit [kPa]

5 Comparison of partitioning approaches

A comparison of T/ET across all seasons for different types of vegetation is presented in Figure 3. The averages were computed over half-hour periods obtained between 6 am and 6 pm local time. For all vegetation types, the average T/ET ratios range from 0.5 in winter to 0.7 in summer. Except for the results over tundra locations, there is a notable similarity across the five methods, particularly during spring and fall. In contrast, the dissimilarity across methods for tundra is possibly linked to the heterogeneity of these sites, which breaks down the assumptions of the partitioning methods more often. A surprising result is that good agreement across methods was observed regardless of measurement height. Even above agricultural sites, where the ratio of z/h can be as large as 30, good agreement was observed for many locations.

Given the underlying uncertainties with partitioning algorithms, one route for verifying the reliability of partitioning estimates is to compare different methods. In Table 5, we compare the different transpiration estimates obtained by CEC, FVS, CEA, and CECw, where the slope α of the linear regression is shown. Results for MREA were similar to those for CEC and are not shown. To help interpret these results, they were separated into four groups: agreement in the 10% range, i.e., $0.9 < \alpha < 1.1$; slopes in the range $0.8 < \alpha < 1.2$; slopes in the range $0.7 < \alpha < 1.3$; disagreement greater than 30%, i.e., $\alpha < 0.7$ or $\alpha > 1.3$.

Overall, T_{FVS} and T_{CEA} are in very good agreement, with differences often within the 10% range across various sites. Only at two sites (PUUM and BARR) were the slopes found to be greater than 1.3 (1.55 and 1.39, respectively). A similar good agreement is seen for the pair T_{FVS} and T_{CECw}, as well as T_{CEA} and T_{CECw}. The close agreement for these three methods was also observed in their average diurnal cycles, in particular between 9 and 15 local time (Figure A1) and even under conditions when $z/h \gg 3$. Nonetheless, note that the diurnal cycles for the FVS exhibit higher values early in the morning and late in the afternoon. Previous studies have highlighted significant variability among the five W options, with differences that can span

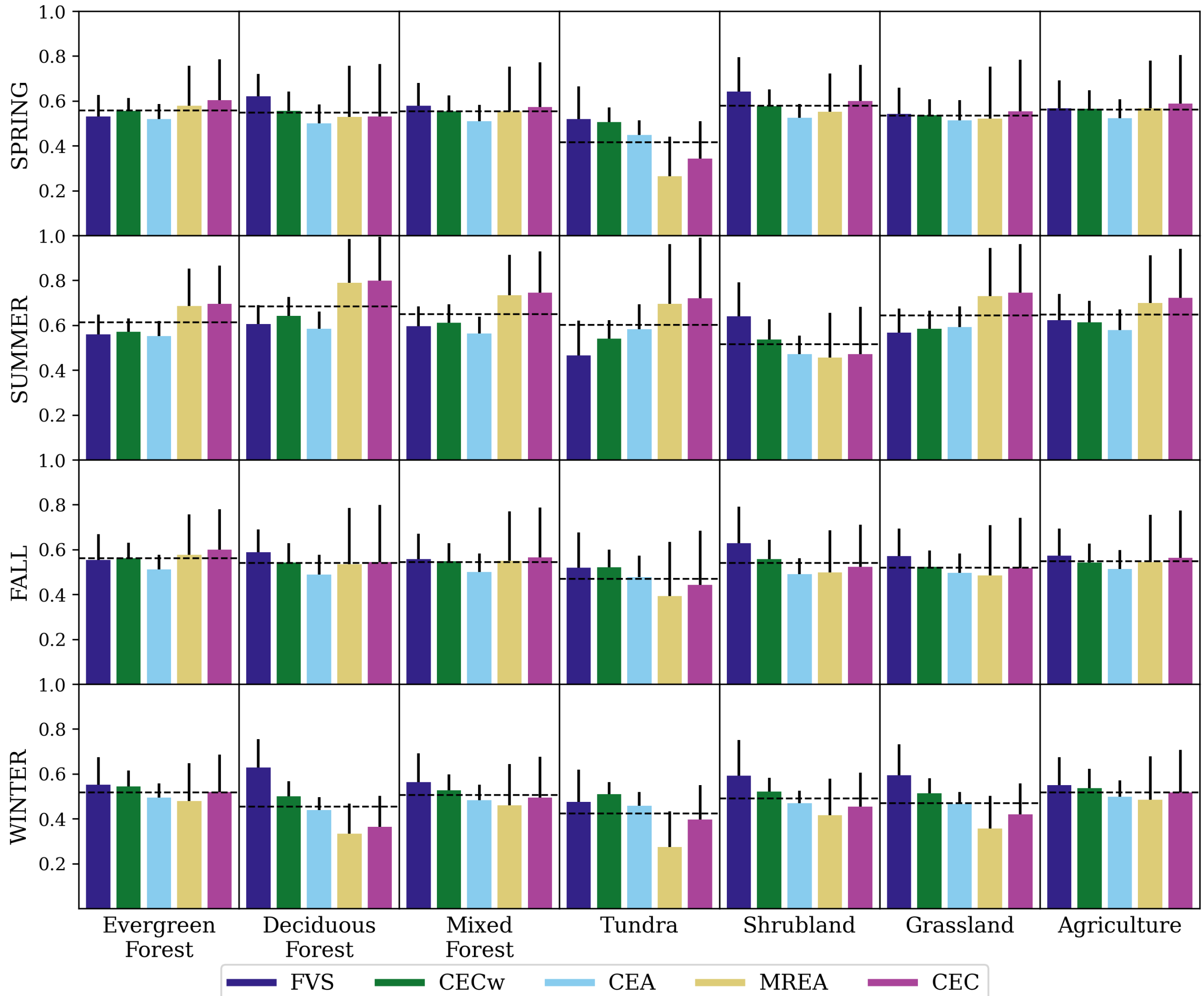

Figure 3. T/ET ratios averaged over all the half-hour periods across seasons and for sites with similar vegetation types. Only half-hour periods during the day (local 6 am and 6 pm) were included, and only when solutions for all methods were available. For FVS and CECw, the ensemble average of outputs with different water-use efficiencies were used. For each season/vegetation type, the black dashed line indicates the average across all methods.

two orders of magnitude during the early hours (Wagle et al., 2023), which inevitably influence the output of the FVS method. Therefore, without additional information on actual water-use efficiency, we cannot explain this behavior with certainty.

The good agreement between FVS and CEA is particularly encouraging given that the CEA method does not require water-use efficiency as an input, using only turbulence information. Moreover, we identified that the outputs of FVS using W_{CR} and W_{CR} were the ones that differed the most from CEA, the remaining three FVS outputs being more similar; W_{OPT} produced the estimates closest to CEA (see Figure A2). Finally, transpiration estimates obtained by CEC tend to disagree more often with other methods (similar trends were observed for MREA, data not shown). Nonetheless, increased agreement is observed when the data are separated by seasons, as shown in Figures 3 and A1.

The main challenge of flux partitioning is verifying the accuracy of these results, so we cannot initially assume that CEC's results are inherently inaccurate or incorrect. However, previous numerical experiments (Zahn et al., 2024) identified that CEC and MREA are more restrictive than the other approaches, being stricter in terms of measurement height and canopy density. Thus, it is possible that the denser canopy during summer directly affect these assumptions, decreasing these two method's performance. They were also more prone to larger errors when the ratios T/E and P/R were more dissimilar. Unfortunately, a metric to determine the "goodness" of partitioning is not feasible since in experiments the true answer is never known. Only estimates of these partitioned fluxes are available since even sap flow, isotope, and flux chamber methods have their limitations.

Given this uncertainty, a prudent recommendation is to implement the ensemble average of T and E of all methods in future studies. Ideally, the results should converge regardless of the partitioning methods used. However, depending on the specific NEON site or analyses of interest, it may be advantageous to prioritize FVS, CEA, and CECw methods under certain circumstances given their close agreement, potentially adding either CEC or MREA since these two methods give similar results and should thus be treated as one member of the ensemble. To this end, the slopes shown in Table 5 offer a reference.

Table 5: Slopes of the linear regression for transpiration for all methods across all sites. Only half-hour periods between 6am and 6pm, without gap filling, were included. For FVS and CECw, their ensemble average across water-use efficiency options was used. Sites were grouped according to the dominant vegetation type representing the tower footprint. To facilitate visual inspection of results, slopes α in the 10% range ($0.9 < \alpha < 1.1$) are shown in bold, while slopes in the 20% range ($0.8 < \alpha < 1.2$) appear in italic.

site	$T_{\text{FVS}}, T_{\text{CEC}}$	$T_{\text{FVS}}, T_{\text{CEA}}$	$T_{\text{FVS}}, T_{\text{CECw}}$	$T_{\text{CECw}}, T_{\text{CEA}}$	$T_{\text{CECw}}, T_{\text{CEC}}$	$T_{\text{CEC}}, T_{\text{CEA}}$
			Evergreen Forest			
ABBY	*1.20*	*0.85*	*0.87*	**1.01**	1.43	0.70
DEJU	*1.16*	*0.86*	*0.82*	**1.07**	1.49	0.72
DELA	1.37	**0.94**	**1.01**	**0.97**	1.42	0.68
GUAN	1.36	**0.92**	**1.08**	**0.93**	1.35	0.68
HEAL	1.26	**0.95**	*0.81*	*1.18*	1.61	0.73
OSBS	1.51	**1.04**	*1.13*	**0.97**	1.43	0.69

PUUM	2.17	1.55	1.39	*1.18*	1.65	0.71
RMNP	1.33	**0.97**	**0.95**	**1.04**	1.44	0.72
SOAP	**1.03**	0.77	0.79	**0.94**	1.34	0.67
TEAK	1.31	**0.91**	**1.04**	*0.90*	1.34	0.67
WREF	**0.91**	0.73	*0.84*	**0.93**	1.24	0.70
YELL	1.27	**0.96**	**0.92**	**1.08**	1.47	0.73
			Deciduous Forest			
CLBJ	1.31	**0.94**	**1.01**	**0.98**	1.35	0.72
GRSM	1.38	**0.94**	**1.03**	**0.94**	1.37	0.69
HARV	1.31	**0.91**	**1.04**	**0.91**	1.28	0.70
LENO	1.32	*0.89*	**1.02**	**0.93**	1.33	0.68
MLBS	1.56	*1.12*	**1.09**	**1.06**	1.48	0.71
ORNL	*1.13*	0.78	*0.89*	*0.89*	1.29	0.67
SCBI	1.33	**0.93**	**1.01**	**0.96**	1.37	0.70
SERC	1.40	**0.96**	**1.09**	**0.91**	1.33	0.69
STEI	1.30	**0.96**	**0.99**	**1.02**	1.37	0.74
UKFS	1.27	**0.91**	**0.98**	**0.97**	1.36	0.72
			Mixed Forest			
BART	1.34	**0.95**	**1.02**	**0.96**	1.34	0.71
BONA	1.26	**0.92**	*0.89*	**1.07**	1.51	0.71
JERC	1.42	**0.97**	**1.07**	**0.96**	1.39	0.69
SJER	1.22	*0.84*	**1.01**	*0.89*	1.34	0.66
TALL	1.24	*0.87*	**0.97**	**0.95**	1.37	0.70
TREE	1.34	**0.94**	**0.98**	**1.01**	1.40	0.71
UNDE	1.39	**0.99**	**1.03**	**1.00**	1.39	0.72
			Tundra			
BARR	1.73	1.51	1.42	**1.10**	1.48	0.74
NIWO	1.36	**1.08**	**0.98**	*1.15*	1.61	0.72
TOOL	1.42	*1.14*	**1.02**	*1.14*	1.53	0.75
			Shrubland			
JORN	**1.05**	0.79	*0.87*	**0.92**	1.24	0.69
MOAB	**1.00**	0.74	*0.81*	*0.90*	1.22	0.69
ONAQ	*1.10*	*0.85*	*0.81*	**1.06**	1.50	0.70

SRER	**1.02**	0.79	*0.87*	**0.92**	1.23	0.70
Grassland						
CPER	**1.08**	*0.85*	**0.90**	**0.96**	1.32	0.70
DCFS	1.50	*1.11*	**1.04**	*1.13*	1.58	0.73
KONZ	1.31	**0.94**	**1.02**	**0.97**	1.38	0.71
NOGP	1.34	**0.98**	**0.94**	**1.08**	1.54	0.72
OAES	1.39	**1.00**	**0.99**	**1.04**	1.46	0.72
WOOD	1.63	*1.18*	*1.14*	**1.07**	1.50	0.72
Agriculture						
BLAN	*1.16*	*0.85*	*0.89*	**1.00**	1.37	0.73
DSNY	1.32	**0.93**	**0.96**	**1.04**	1.49	0.70
KONA	**1.04**	*0.81*	*0.89*	**0.94**	1.24	0.72
LAJA	1.28	**0.93**	**1.06**	**0.92**	1.27	0.72
STER	**0.94**	0.75	*0.87*	*0.87*	*1.12*	0.72

Table 6: Slopes of the linear regression for net primary productivity (P) for all methods across all sites. Only half-hour periods between 6 am and 6 pm (local time) were included. For FVS and CECw, their ensemble average across water-use efficiency options was used. Sites were grouped according to the dominant vegetation type. To facilitate visual inspection of results, slopes α in the 10% range ($0.9 < \alpha < 1.1$) are shown in bold, while slopes in the 20% range ($0.8 < \alpha < 1.2$) appear in italic.

site	$P_{\mathrm{FVS}}, P_{\mathrm{CEC}}$	$P_{\mathrm{FVS}}, P_{\mathrm{CEA}}$	$P_{\mathrm{FVS}}, P_{\mathrm{CECw}}$	$P_{\mathrm{CECw}}, P_{\mathrm{CEA}}$	$P_{\mathrm{CECw}}, P_{\mathrm{CEC}}$	$P_{\mathrm{CEC}}, P_{\mathrm{CEA}}$
Evergreen Forest						
ABBY	0.32	0.57	*0.87*	0.32	0.18	1.64
DEJU	0.09	0.16	0.73	0.07	0.04	1.56
DELA	0.45	*0.81*	**0.95**	0.67	0.37	1.52
GUAN	0.53	0.69	**1.01**	0.29	0.20	1.67
HEAL	0.15	0.21	0.75	0.13	0.08	1.43
OSBS	0.36	0.68	**1.05**	0.36	0.17	1.74
PUUM	0.60	**1.01**	1.48	0.14	0.08	1.60
RMNP	0.14	0.21	*0.87*	0.13	0.08	1.49
SOAP	0.11	0.23	*0.80*	0.10	0.05	1.69
TEAK	0.30	0.46	**0.97**	0.24	0.13	1.84
WREF	0.18	0.34	*0.82*	0.16	0.10	1.57

YELL	0.12	0.24	*0.86*	0.12	0.07	1.53
			Deciduous Forest			
CLBJ	0.36	0.61	**0.92**	0.61	0.38	1.45
GRSM	0.53	**0.91**	**1.09**	0.57	0.32	1.62
HARV	0.62	**1.07**	**1.06**	0.72	0.40	1.59
LENO	0.49	**0.92**	**0.97**	*0.83*	0.44	1.61
MLBS	0.57	**0.91**	*1.11*	0.44	0.27	1.47
ORNL	0.47	**0.92**	*0.88*	0.80	0.39	1.73
SCBI	0.49	*0.85*	**0.97**	0.77	0.44	1.52
SERC	0.60	**1.00**	**1.06**	0.79	0.46	1.52
STEI	0.32	0.53	**0.95**	0.40	0.23	1.43
UKFS	0.34	0.58	*0.88*	0.45	0.26	1.50
			Mixed Forest			
BART	0.52	**0.92**	**1.03**	0.52	0.27	1.58
BONA	0.13	0.25	*0.80*	0.12	0.06	1.56
JERC	0.30	0.59	**0.97**	0.52	0.26	1.61
SJER	0.33	0.53	**1.04**	0.34	0.20	1.74
TALL	0.38	0.65	**0.94**	0.43	0.23	1.64
TREE	0.40	0.70	**0.94**	0.55	0.30	1.53
UNDE	0.38	0.67	**0.97**	0.41	0.22	1.49
			Tundra			
BARR	0.05	0.10	1.20	0.04	0.02	1.54
NIWO	0.03	0.07	*0.84*	0.04	0.02	1.48
TOOL	0.05	0.12	**0.99**	0.07	0.03	1.42
			Shrubland			
JORN	0.08	0.17	*0.84*	0.06	0.03	1.81
MOAB	0.01	0.03	*0.90*	-0.00	-0.00	1.75
ONAQ	0.04	0.10	*0.88*	0.05	0.02	1.54
SRER	0.10	0.22	*0.86*	0.12	0.06	1.84
			Grassland			
CPER	0.06	0.17	**0.94**	0.08	0.04	1.56
DCFS	0.13	0.22	**0.92**	0.14	0.09	1.48
KONZ	0.33	0.60	**0.99**	0.41	0.22	1.51

NOGP	0.17	0.29	**0.98**	0.17	0.10	1.61
OAES	0.21	0.36	**0.95**	0.25	0.15	1.48
WOOD	0.22	0.39	**1.09**	0.22	0.12	1.52
			Agriculture			
BLAN	0.39	0.67	**0.92**	0.44	0.25	1.54
DSNY	0.35	0.72	**0.98**	0.34	0.16	1.83
KONA	0.05	0.17	**0.91**	0.09	0.03	1.50
LAJA	0.43	0.79	**0.96**	0.71	0.40	1.48
STER	0.05	0.15	*0.88*	0.05	0.02	1.58

In contrast to ET flux partitioning, the estimated CO_2 flux components showed more disagreement across methods (Table 6). Generally, methods that do not require water-use efficiency (CEC, CEA, and MREA) underestimated P and R compared to FVS and CECw, where most slopes indicated less than 10% of agreement. Our previous findings using large-eddy simulations to test all methods (Zahn et al., 2024) found larger errors for CO_2 components by CEC, CEA, and MREA. We identified that while CO_2 is essential as a tracer for H_2O in these methods' formulations, it cannot be reliably partitioned between soil and canopy components given the lack of an upper bound on Fc components, which combine a positive (respiration) and a negative flux (photosynthesis). In contrast, FVS and CECw limit the magnitude of these components by imposing $W = P/T$, thus adding a constraint to the magnitude of $Fc = R + P$. Note, however, that P (and R) will be subjected to uncertainties present in the water-use efficiency estimate. Overall, these findings indicate that CO_2 flux partitioning is more challenging, and care must be taken when using and interpreting these results. At present, for CO_2 flux components we thus recommend taking the ensemble average of CECw and FVS (across all five W estimates), potentially adding CEA as a third member of the ensemble for select sites, in particular above deciduous forests, following Table 6.

6 Exploring research opportunities with a flux partitioning dataset

The final dataset will enable hydrologists and ecologists to investigate a variety of research questions related to flux partitioning. Considering the dataset's breadth of coverage, its spatial and temporal distributions enable comparison across diverse conditions, encompassing various forest structures, water availability, climate and weather patterns, and more. Due to remaining gaps in the final flux time series, the evaluation of long-term budgets over the period of five years is not recommended without additional work to complete the temporal coverage. Nonetheless, budgets over smaller periods and for specific NEON sites are possible and can be explored. In this case, for nighttime periods, when most partitioning results are unavailable, researchers might fill gaps by assuming zero plant fluxes (for example for T and P), and $E = ET$ and $R = F_c$.

One relevant research question regarding ecosystem functioning revolves around the main drivers of plant and soil fluxes. To briefly explore this question, we implemented XGBoost to estimate the importance of a few environmental variables on T and T/ET, where we first averaged the output from all five partitioning models (excluding gap-filled periods). Note that this analysis is different from the procedure used for gap filling, where ET was used as a predictor. Here, we simply investigate the importance of each environmental variable as a driver of plant fluxes, and thus ET is not used as an input. The selected features were air temperature, photosynthetically active radiation, relative humidity, vapor pressure deficit, mean wind speed, sensible heat flux, and friction velocity.

The importance of each environmental variable is shown in Figure 4 for T (top row) and T/ET (bottom row) averaged across sites with the same vegetation type. The standard deviation was also computed across sites and is represented by the bars in the figure (the outer circle represents a standard deviation of 20%). Notably, these seven environmental variables can explain up to 77% ($R^2 = 0.77$ over agricultural sites) of the variability in transpiration. With exception of tundra sites, where sensible heat was the most important variable, photosynthetically active radiation, followed by air temperature, were the most relevant features. Intersite variability was also observed; for instance, the standard deviation of PAR and T_{AIR} was as large as 20% for evergreen and mixed forests. The variability across methods is not included in this analysis since their ensemble average was used; nonetheless, similar to our discussion regarding gap filling, we observed that FVS combined with W_{CC} and W_{CR} often differed from the other methods (results not shown). This discrepancy is also left for future studies.

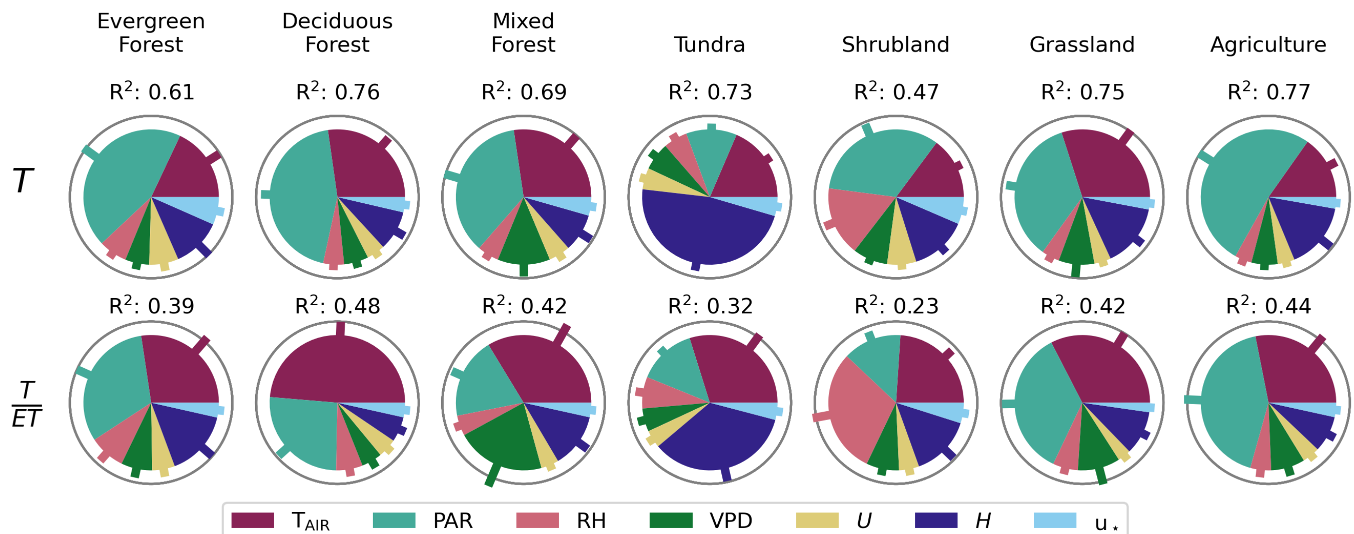

Figure 4. Importance of seven environmental variables as drivers of transpiration (top row) and T/ET (bottom row) across vegetation types. From left to right, the legend includes air temperature (T_{AIR}), photosynthetically active radiation (PAR), relative humidity (RH), vapor pressure deficit (VPD), mean wind velocity (U), sensible heat flux (H), and friction velocity (u_*). Pie charts display the average over all sites with the same vegetation type, where bars represent the standard deviation of feature importance across sites (in percentage). As a reference, the outer circle represents a standard deviation of 20%.

The ratio T/ET, on the other hand, is more complex and dependent on additional variables not include in this analysis. Across vegetation types, these seven features were able to explain as little as 23% (shrublands), to no more than 48% (deciduous forests). While relevant environmental variables such as soil moisture and accumulated rainfall were not included in our analyses due to availability and/or data quality, these variables can be included for selected sites and/or smaller temporal records. By exploring different environmental forcings, this dataset presents an opportunity to address specific inquiries, such as the impact of heat waves, droughts, and other extreme events on plant transpiration. Research endeavors of this nature are vital, as they contribute to a deeper comprehension of how present and future climate conditions affect local water and energy budgets, as well as plant water-use efficiency (Hatfield and Dold, 2019).

The link between fluxes and canopy structure features another set of critical scientific questions that can be addressed using this dataset. Previous work (Wang et al., 2014; Wei et al., 2017) compiled results from several case studies and summarized the correlation between T/ET and leaf-area-index (LAI), which has been turned into a simple regression model often used in other studies (Nelson et al., 2020). Future investigations could expand this study, and utilize NEON products or satellite imagery to derive seasonal variations in LAI, thereby enhancing our understanding of how vegetation dynamics influence transpiration. By combining the most important meteorological features and vegetation characteristics, simpler and more accurate partitioning models can be derived and more broadly applied.

In addition to observational studies, modeling or hybrid studies (Rafi et al., 2019; Kozii et al., 2020) can also benefit from the dataset compiled here. This approach can be explored to develop models for transpiration partitioning or modeling (Bright et al., 2022), where observations can validate the model or help determine relevant parameters, such as plant conductance. For instance, Schreiner-McGraw et al. (2022) used transpiration data from FVS to optimize parameters for ParFlow.CLM, a hydrological model coupling groundwater flow to a land-surface model (Maxwell and Miller, 2005). This methodology is particularly useful for evaluating the performance of land surface models, such as NOAH-MP (Niu et al., 2011; Li et al., 2021) and the Community Land Model (Lawrence et al., 2007), thereby improving assessments of their impact on weather prediction and climate projections (Berg and Sheffield, 2019; Dong et al., 2022).

Future studies can also expand this dataset by including additional partitioning models, including the products derived from the CO_2 partitioning algorithm proposed by Reichstein et al. (2005) and Lasslop et al. (2010). In terms of ET partitioning, for instance, different methods have been proposed that take GPP as an input (Perez-Priego et al., 2018; Zhou et al., 2016; Nelson et al., 2018); such models have been compared by Nelson et al. (2020) across FLUXNET sites . By comparing different algorithms, we can further explore their uncertainties and focus on model improvement. Finally, as more data become available, other options can be used to train machine learning algorithms focusing on gap filling.

7 Conclusions

This paper described the first dataset of flux partitioning across all NEON eddy-covariance sites from 2019 to 2023. Estimates of transpiration, evaporation, soil respiration and plant net CO_2 assimilation were obtained by five partitioning approaches. These methods had undergone extensive testing in both experimental and numerical settings in prior studies (Zahn et al., 2022, 2024), aiding in understanding their limitations and reliability

Overall, ET components exhibited notable consistency across methods, particularly evident in FVS, CEA, and CECw, thereby enhancing confidence in these estimates. In contrast, CO_2 components are more challenging, and care must be taken when interpreting these estimates. The final dataset, available at https://doi.org/10.5281/zenodo.12191876, can be used to address numerous research questions. We recommend utilizing an ensemble average of methods available for a particular site. However, based on the comparison presented in this study, a combination of methods might be adopted for specific analyses or NEON sites. In any case, the user should keep in mind the following considerations:

1. None of the methods can estimate plant respiration and gross photosynthesis separately; only an estimate of their difference, plant net photosynthesis (i.e., plant net fluxes) is computed. Similarly, below canopy respiration — which we assume to be mostly from the soil — also includes respiration from roots and litter;
2. Evaporation from canopy intercepted water is neglected;
3. While nighttime fluxes are also included, we do not recommend the inclusion of nighttime periods since CO_2 assimilation is not expected at night for C3 and C4 plants, thus breaking the assumption of CO_2/H_2O coupling;
4. Uncertainties in W will be transferred to the partitioning estimates for FVS and CECw;
5. Given the differences in the methods' frameworks and their assumptions, data gaps vary. Thus, selecting the periods of interest and methods to average must be done with care to avoid biases.

This dataset opens many research pathways that we outlined above, and can be the basis for other derived datasets. The partitioning approaches can also be applied to other periods of NEON observations, or implemented in the NEON workflow so that the partitioned fluxes are part of the standard NEON outputs.

8 Code and data availability

The dataset is available at https://doi.org/10.5281/zenodo.12191876 (Zahn and Bou-Zeid, 2024). In addition to all flux components, it also contains the auxiliary meteorological inputs used to implement the Extreme Gradient Boosting algorithm for gap filling and feature importance analysis. The scripts used to implement all five partitioning methods can be found at https://github.com/einaraz/PartitioningMethods (Zahn, 2024).

Appendix A: Diurnal cycle of T/ET

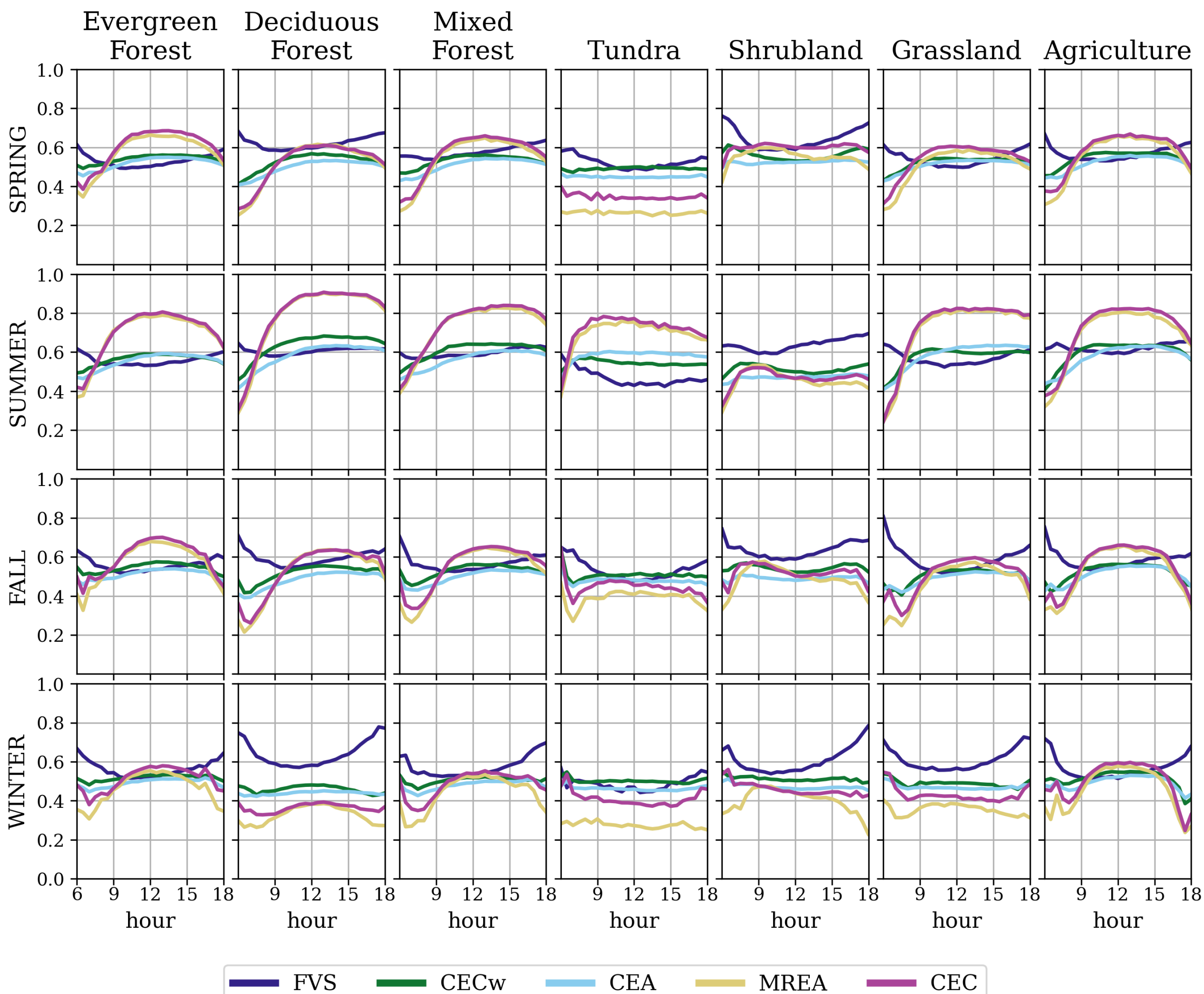

Figure A1. Diurnal cycle of the partitioning ratios T/ET for all methods across seasons and vegetation type. Half-hourly estimates during daytime hours (local 6 am to 6 pm) are only included when all methods have valid solutions. For FVS and CECw, their ensemble average across water-use efficiency options was used.

Author contributions. EZ and EBZ conceptualised the study and designed the methodology; EZ processed the data; EZ wrote the original draft; EZ and EBZ edited and revised the manuscript.

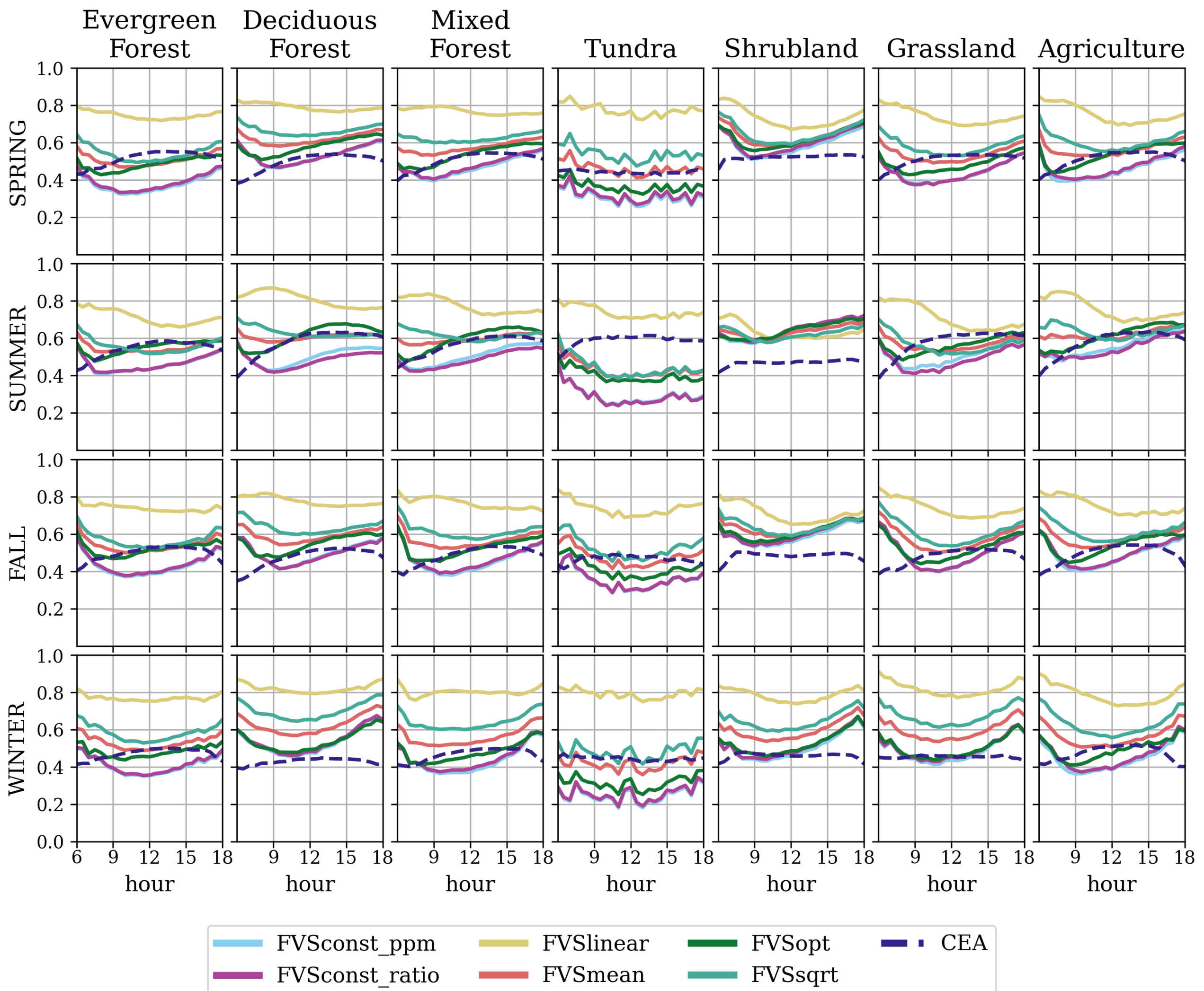

Figure A2. Diurnal cycle of the partitioning ratios T/ET comparing the FVS output from different water-use efficiency models (and their ensemble average) and the CEA method (dahsed line). Half-hourly estimates during daytime hours (local 6 am to 6 pm) are only included when all methods have valid solutions.

Competing interests. The authors declare that there are no competing interests

Acknowledgements. This work was supported by the Moore Charitable Foundation Science-to-Action Fund from the School of Engineering and Applied Science at Princeton University.

References

Baslam, M., Mitsui, T., Hodges, M., Priesack, E., Herritt, M. T., Aranjuelo, I., and Sanz-Saez, A.: Photosynthesis in a Changing Global Climate: Scaling Up and Scaling Down in Crops, Frontiers in Plant Science, 11, https://doi.org/10.3389/fpls.2020.00882, 2020.

Berg, A. and Sheffield, J.: Evapotranspiration Partitioning in CMIP5 Models: Uncertainties and Future Projections, Journal of Climate, 32, 2653–2671, https://doi.org/10.1175/jcli-d-18-0583.1, 2019.

Bright, R. M., Miralles, D. G., Poyatos, R., and Eisner, S.: Simple Models Outperform More Complex Big-Leaf Models of Daily Transpiration in Forested Biomes, Geophysical Research Letters, 49, https://doi.org/10.1029/2022gl100100, 2022.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, ACM, https://doi.org/10.1145/2939672.2939785, 2016.

Dong, J., Lei, F., and Crow, W. T.: Land transpiration-evaporation partitioning errors responsible for modeled summertime warm bias in the central United States, Nature Communications, 13, https://doi.org/10.1038/s41467-021-27938-6, 2022.

Dusenge, M. E., Duarte, A. G., and Way, D. A.: Plant carbon metabolism and climate change: elevated CO2 and temperature impacts on photosynthesis, photorespiration and respiration, New Phytologist, 221, 32–49, https://doi.org/https://doi.org/10.1111/nph.15283, 2019.

Fiorella, R. P., Good, S. P., Allen, S. T., Guo, J. S., Still, C. J., Noone, D. C., Anderegg, W. R. L., Florian, C. R., Luo, H., Pingintha-Durden, N., and Bowen, G. J.: Calibration Strategies for Detecting Macroscale Patterns in NEON Atmospheric Carbon Isotope Observations, Journal of Geophysical Research: Biogeosciences, 126, https://doi.org/10.1029/2020jg005862, 2021.

Good, S. P., Soderberg, K., Guan, K., King, E. G., Scanlon, T. M., and Caylor, K. K.: δ2H isotopic flux partitioning of evapotranspiration over a grass field following a water pulse and subsequent dry down, Water Resources Research, 50, 1410–1432, https://doi.org/10.1002/2013wr014333, 2014.

Grinsztajn, L., Oyallon, E., and Varoquaux, G.: Why do tree-based models still outperform deep learning on typical tabular data?, in: Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track, https://openreview.net/forum?id=Fp7__phQszn, 2022.

Hatfield, J. L. and Dold, C.: Water-Use Efficiency: Advances and Challenges in a Changing Climate, Frontiers in Plant Science, 10, https://doi.org/10.3389/fpls.2019.00103, 2019.

Huang, I.-H. and Hsieh, C.-I.: Gap-Filling of Surface Fluxes Using Machine Learning Algorithms in Various Ecosystems, Water, 12, 3415, https://doi.org/10.3390/w12123415, 2020.

Irvin, J., Zhou, S., McNicol, G., Lu, F., Liu, V., Fluet-Chouinard, E., Ouyang, Z., Knox, S. H., Lucas-Moffat, A., Trotta, C., Papale, D., Vitale, D., Mammarella, I., Alekseychik, P., Aurela, M., Avati, A., Baldocchi, D., Bansal, S., Bohrer, G., Campbell, D. I., Chen, J., Chu, H., Dalmagro, H. J., Delwiche, K. B., Desai, A. R., Euskirchen, E., Feron, S., Goeckede, M., Heimann, M., Helbig, M., Helfter, C., Hemes, K. S., Hirano, T., Iwata, H., Jurasinski, G., Kalhori, A., Kondrich, A., Lai, D. Y., Lohila, A., Malhotra, A., Merbold, L., Mitra, B., Ng, A., Nilsson, M. B., Noormets, A., Peichl, M., Rey-Sanchez, A. C., Richardson, A. D., Runkle, B. R., Schäfer, K. V., Sonnentag, O., Stuart-Haëntjens, E., Sturtevant, C., Ueyama, M., Valach, A. C., Vargas, R., Vourlitis, G. L., Ward, E. J., Wong, G. X., Zona, D., Alberto, M. C. R., Billesbach, D. P., Celis, G., Dolman, H., Friborg, T., Fuchs, K., Gogo, S., Gondwe, M. J., Goodrich, J. P., Gottschalk, P., Hörtnagl, L., Jacotot, A., Koebsch, F., Kasak, K., Maier, R., Morin, T. H., Nemitz, E., Oechel, W. C., Oikawa, P. Y., Ono, K., Sachs, T., Sakabe, A., Schuur, E. A., Shortt, R., Sullivan, R. C., Szutu, D. J., Tuittila, E.-S., Varlagin, A., Verfaillie, J. G., Wille, C., Windham-Myers, L., Poulter, B., and Jackson, R. B.: Gap-filling eddy covariance methane fluxes: Comparison of machine learning model predictions and uncertainties at FLUXNET-CH4 wetlands, Agricultural and Forest Meteorology, 308-309, 108 528, https://doi.org/https://doi.org/10.1016/j.agrformet.2021.108528, 2021.

Kirschbaum, M. U. F. and McMillan, A. M. S.: Warming and Elevated CO_2 Have Opposing Influences on Transpiration. Which is more Important?, Current Forestry Reports, 4, 51–71, https://doi.org/10.1007/s40725-018-0073-8, 2018.

Klosterhalfen, A., Graf, A., Brüggemann, N., Drüe, C., Esser, O., González-Dugo, M. P., Heinemann, G., Jacobs, C. M. J., Mauder, M., Moene, A. F., Ney, P., Pütz, T., Rebmann, C., Ramos Rodríguez, M., Scanlon, T. M., Schmidt, M., Steinbrecher, R., Thomas, C. K., Valler, V., Zeeman, M. J., and Vereecken, H.: Source partitioning of H_2O and CO_2 fluxes based on high-frequency eddy covariance data: a comparison between study sites, Biogeosciences, 16, 1111–1132, https://doi.org/10.5194/bg-16-1111-2019, 2019.

Kool, D., Agam, N., Lazarovitch, N., Heitman, J., Sauer, T., and Ben-Gal, A.: A review of approaches for evapotranspiration partitioning, Agricultural and Forest Meteorology, 184, 56–70, https://doi.org/10.1016/j.agrformet.2013.09.003, 2014.

Kozii, N., Haahti, K., Tor-ngern, P., Chi, J., Hasselquist, E. M., Laudon, H., Launiainen, S., Oren, R., Peichl, M., Wallerman, J., and Hasselquist, N. J.: Partitioning growing season water balance within a forested boreal catchment using sap flux, eddy covariance, and a process-based model, Hydrology and Earth System Sciences, 24, 2999–3014, https://doi.org/10.5194/hess-24-2999-2020, 2020.

Lasslop, G., Reichstein, M., Papale, D., Richardson, A. D., Arneth, A., Barr, A., Stoy, P., and Wohlfahrt, G.: Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation, Global Change Biology, 16, 187–208, https://doi.org/https://doi.org/10.1111/j.1365-2486.2009.02041.x, 2010.

Lawrence, D. M., Thornton, P. E., Oleson, K. W., and Bonan, G. B.: The Partitioning of Evapotranspiration into Transpiration, Soil Evaporation, and Canopy Evaporation in a GCM: Impacts on Land–Atmosphere Interaction, Journal of Hydrometeorology, 8, 862–880, https://doi.org/10.1175/jhm596.1, 2007.

Lesk, C., Coffel, E., Winter, J., Ray, D., Zscheischler, J., Seneviratne, S. I., and Horton, R.: Stronger temperature–moisture couplings exacerbate the impact of climate warming on global crop yields, Nature Food, 2, 683–691, https://doi.org/10.1038/s43016-021-00341-6, 2021.

Li, L., Yang, Z., Matheny, A. M., Zheng, H., Swenson, S. C., Lawrence, D. M., Barlage, M., Yan, B., McDowell, N. G., and Leung, L. R.: Representation of Plant Hydraulics in the Noah-MP Land Surface Model: Model Development and Multiscale Evaluation, Journal of Advances in Modeling Earth Systems, 13, https://doi.org/10.1029/2020ms002214, 2021.

Lunch, C., Laney, C., Mietkiewicz, N., Sokol, E., Cawley, K., and NEON: neonUtilities: Utilities for Working with NEON Data, https://doi.org/10.32614/CRAN.package.neonUtilities, version 2.4.2, 2024.

Maxwell, R. M. and Condon, L. E.: Connections between groundwater flow and transpiration partitioning, Science, 353, 377–380, https://doi.org/10.1126/science.aaf7891, 2016.

Maxwell, R. M. and Miller, N. L.: Development of a Coupled Land Surface and Groundwater Model, Journal of Hydrometeorology, 6, 233–247, https://doi.org/10.1175/jhm422.1, 2005.

Mengis, N., Keller, D. P., Eby, M., and Oschlies, A.: Uncertainty in the response of transpiration to CO2 and implications for climate change, Environmental Research Letters, 10, 094 001, https://doi.org/10.1088/1748-9326/10/9/094001, 2015.

Metzger, S., Durden, D., Sturtevant, C., Luo, H., Pingintha-Durden, N., Sachs, T., Serafimovich, A., Hartmann, J., Li, J., Xu, K., and Desai, A. R.: eddy4R 0.2.0: a DevOps model for community-extensible processing and analysis of eddy-covariance data based on R, Git, Docker, and HDF5, Geoscientific Model Development, 10, 3189–3206, https://doi.org/10.5194/gmd-10-3189-2017, 2017.

Metzger, S., Ayres, E., Durden, D., Florian, C., Lee, R., Lunch, C., Luo, H., Pingintha-Durden, N., Roberti, J. A., SanClements, M., Sturtevant, C., Xu, K., and Zulueta, R. C.: From NEON Field Sites to Data Portal: A Community Resource for Surface–Atmosphere Research Comes Online, Bulletin of the American Meteorological Society, 100, 2305 – 2325, https://doi.org/10.1175/BAMS-D-17-0307.1, 2019.

Metzger, S., Durden, D., Xu, K., Pingintha-Durden, N., Luo, H., and Florian, C.: NEON Algorithm Theoretical Basis Document (ATBD): Eddy-Covariance Data Products Bundle, Technical Report NEON.DOC.004571, National Ecological Observatory Network (NEON), available from NEON: http://data.neonscience.org/data-product-view?dpCode=DP4.00200.001, 2022.

National Ecological Observatory Network (NEON): Bundled data products - eddy covariance (DP4.00200.001), https://doi.org/10.48443/J9PT-M241, 2024a.

National Ecological Observatory Network (NEON): Bundled data products - eddy covariance (DP4.00200.001), https://data.neonscience.org/data-products/DP4.00200.001, 2024b.

National Ecological Observatory Network (NEON): Photosynthetically active radiation (PAR) (DP1.00024.001), https://data.neonscience.org/data-products/DP1.00024.001, 2024c.

National Ecological Observatory Network (NEON): Photosynthetically active radiation (PAR) (DP1.00024.001), https://doi.org/10.48443/VMJD-YY64, 2024d.

National Ecological Observatory Network (NEON): Relative humidity (DP1.00098.001), https://data.neonscience.org/data-products/DP1.00098.001, 2024e.

National Ecological Observatory Network (NEON): Relative humidity (DP1.00098.001), https://doi.org/10.48443/K9VK-5K27, 2024f.

National Ecological Observatory Network (NEON): Triple aspirated air temperature (DP1.00003.001), https://data.neonscience.org/data-products/DP1.00003.001, 2024g.

National Ecological Observatory Network (NEON): Triple aspirated air temperature (DP1.00003.001), https://doi.org/10.48443/PDE7-K607, 2024h.

National Ecological Observatory Network (NEON): Shortwave and longwave radiation (net radiometer) (DP1.00023.001), https://data.neonscience.org/data-products/DP1.00023.001, 2024i.

National Ecological Observatory Network (NEON): Shortwave and longwave radiation (net radiometer) (DP1.00023.001), https://doi.org/10.48443/9QPC-5V70, 2024j.

National Ecological Observatory Network (NEON): 2D wind speed and direction (DP1.00001.001), https://doi.org/10.48443/YEX7-7Z81, 2024k.

National Ecological Observatory Network (NEON): 2D wind speed and direction (DP1.00001.001), https://data.neonscience.org/data-products/DP1.00001.001, 2024l.

Nelson, J. A., Carvalhais, N., Cuntz, M., Delpierre, N., Knauer, J., Ogée, J., Migliavacca, M., Reichstein, M., and Jung, M.: Coupling Water and Carbon Fluxes to Constrain Estimates of Transpiration: The TEA Algorithm, Journal of Geophysical Research: Biogeosciences, 123, 3617–3632, https://doi.org/10.1029/2018JG004727, 2018.

Nelson, J. A., Pérez-Priego, O., Zhou, S., Poyatos, R., Zhang, Y., Blanken, P. D., Gimeno, T. E., Wohlfahrt, G., Desai, A. R., Gioli, B., Limousin, J.-M., Bonal, D., Paul-Limoges, E., Scott, R. L., Varlagin, A., Fuchs, K., Montagnani, L., Wolf, S., Delpierre, N., Berveiller, D., Gharun, M., Belelli Marchesini, L., Gianelle, D., Šigut, L., Mammarella, I., Siebicke, L., Andrew Black, T., Knohl, A., Hörtnagl, L., Magliulo, V., Besnard, S., Weber, U., Carvalhais, N., Migliavacca, M., Reichstein, M., and Jung, M.: Ecosystem transpiration and evaporation: Insights from three water flux partitioning methods across FLUXNET sites, Global Change Biology, n/a, https://doi.org/10.1111/gcb.15314, 2020.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, Journal of Geophysical Research, 116, https://doi.org/10.1029/2010jd015139, 2011.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Reichstein, M., Ribeca, A., van Ingen, C., Vuichard, N., Zhang, L., Amiro, B., Ammann, C., Arain, M. A., Ardö, J., Arkebauer, T., Arndt, S. K., Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L. B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Boike, J., Bolstad, P. V., Bonal, D., Bonnefond, J.-M., Bowling, D. R., Bracho, R., Brodeur, J., Brümmer, C., Buchmann, N., Burban, B., Burns, S. P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T. R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B. D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P. S., D'Andrea, E., da Rocha, H., Dai, X., Davis, K. J., Cinti, B. D., Grandcourt, A. d., Ligne, A. D., De Oliveira, R. C., Delpierre, N., Desai, A. R., Di Bella, C. M., Tommasi, P. d., Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrêne, E., Dunn, A., Dušek, J., Eamus, D., Eichelmann, U., ElKhidir, H. A. M., Eugster, W., Ewenz, C. M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., Gharun, M., Gianelle, D., Gielen, B., Gioli, B., Gitelson, A., Goded, I., Goeckede, M., Goldstein, A. H., Gough, C. M., Goulden, M. L., Graf, A., Griebel, A., Gruening, C., Grünwald, T., Hammerle, A., Han, S., Han, X., Hansen, B. U., Hanson, C., Hatakka, J., He, Y., Hehn, M., Heinesch, B., Hinko-Najera, N., Hörtnagl, L., Hutley, L., Ibrom, A., Ikawa, H., Jackowicz-Korczynski, M., Janouš, D., Jans, W., Jassal, R., Jiang, S., Kato, T., Khomik, M., Klatt, J., Knohl, A., Knox, S., Kobayashi, H., Koerber, G., Kolle, O., Kosugi, Y., Kotani, A., Kowalski, A., Kruijt, B., Kurbatova, J., Kutsch, W. L., Kwon, H., Launiainen, S., Laurila, T., Law, B., Leuning, R., Li, Y., Liddell, M., Limousin, J.-M., Lion, M., Liska, A. J., Lohila, A., López-Ballesteros, A., López-Blanco, E., Loubet, B., Loustau, D., Lucas-Moffat, A., Lüers, J., Ma, S., Macfarlane, C., Magliulo, V., Maier, R., Mammarella, I., Manca, G., Marcolla, B., Margolis, H. A., Marras, S., Massman, W., Mastepanov, M., Matamala, R., Matthes, J. H., Mazzenga, F., McCaughey, H., McHugh, I., McMillan, A. M. S., Merbold, L., Meyer, W., Meyers, T., Miller, S. D., Minerbi, S., Moderow, U., Monson, R. K., Montagnani, L., Moore, C. E., Moors, E., Moreaux, V., Moureaux, C., Munger, J. W., Nakai, T., Neirynck, J., Nesic, Z., Nicolini, G., Noormets, A., Northwood, M., Nosetto, M., Nouvellon, Y., Novick, K., Oechel, W., Olesen, J. E., Ourcival, J.-M., Papuga, S. A., Parmentier, F.-J., Paul-Limoges, E., Pavelka, M., Peichl, M., Pendall, E., Phillips, R. P., Pilegaard, K., Pirk, N., Posse, G., Powell, T., Prasse, H., Prober, S. M., Rambal, S., Rannik, U., Raz-Yaseef, N., Rebmann, C., Reed, D., Dios, V. R. d., Restrepo-Coupe, N., Reverter, B. R., Roland, M., Sabbatini, S., Sachs, T., Saleska, S. R., Sánchez-Cañete, E. P., Sanchez-Mejia, Z. M., Schmid, H. P., Schmidt, M., Schneider, K., Schrader, F., Schroder, I., Scott, R. L., Sedlák, P., Serrano-Ortíz, P., Shao, C., Shi, P., Shironya, I., Siebicke, L., Šigut, L., Silberstein, R., Sirca, C., Spano, D., Steinbrecher, R., Stevens, R. M., Sturtevant, C., Suyker, A., Tagesson, T., Takanashi, S., Tang, Y., Tapper, N., Thom, J., Tomassucci, M., Tuovinen, J.-P., Urbanski, S., Valentini, R., van der Molen, M., van Gorsel, E., van Huissteden, K., Varlagin, A., Verfaillie, J., Vesala, T., Vincke, C., Vitale, D., Vygodskaya, N., Walker, J. P., Walter-Shea, E., Wang, H., Weber, R., Westermann, S., Wille, C., Wofsy, S., Wohlfahrt, G., Wolf, S., Woodgate, W., Li, Y., Zampedri, R., Zhang, J., Zhou, G., Zona, D., Agarwal, D., Biraud, S., Torn, M., and Papale, D.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Scientific Data, 7, https://doi.org/10.1038/s41597-020-0534-3, 2020.

Perez-Priego, O., Katul, G. G., Reichstein, M., El-Madany, T. S., Ahrens, B., Carrara, A., Scanlon, T. M., and Migliavacca, M.: Partitioning Eddy Covariance Water Flux Components Using Physiological and Micrometeorological Approaches, Journal of Geophysical Research: Biogeosciences, 123, 3353–3370, https://doi.org/10.1029/2018JG004637, 2018.

Rafi, Z., Merlin, O., Le Dantec, V., Khabba, S., Mordelet, P., Er-Raki, S., Amazirh, A., Olivera-Guerra, L., Ait Hssaine, B., Simonneaux, V., Ezzahar, J., and Ferrer, F.: Partitioning evapotranspiration of a drip-irrigated wheat crop: Inter-comparing eddy covariance-, sap flow-, lysimeter- and FAO-based methods, Agricultural and Forest Meteorology, 265, 310–326, https://doi.org/10.1016/j.agrformet.2018.11.031, 2019.

Rebmann, C., Kolle, O., Heinesch, B., Queck, R., Ibrom, A., and Aubinet, M.: Data acquisition and flux calculations, in: Eddy Covariance: A Practical Guide to Measurement and Data Analysis, edited by Aubinet, M., Vesala, T., and Papale, D., pp. 59–83, Springer, Dordrecht, Heidelberg, London, New York, 2012.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., and Valentini, R.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, Global Change Biology, 11, 1424–1439, https://doi.org/https://doi.org/10.1111/j.1365-2486.2005.001002.x, 2005.

Scanlon, T. M. and Kustas, W. P.: Partitioning carbon dioxide and water vapor fluxes using correlation analysis, Agricultural and Forest Meteorology, 150, 89 – 99, https://doi.org/https://doi.org/10.1016/j.agrformet.2009.09.005, 2010.

Scanlon, T. M. and Sahu, P.: On the correlation structure of water vapor and carbon dioxide in the atmospheric surface layer: A basis for flux partitioning, Water Resources Research, 44, https://doi.org/10.1029/2008WR006932, 2008.

Scanlon, T. M., Schmidt, D. F., and Skaggs, T. H.: Correlation-based flux partitioning of water vapor and carbon dioxide fluxes: Method simplification and estimation of canopy water use efficiency, Agricultural and Forest Meteorology, 279, 107 732, https://doi.org/https://doi.org/10.1016/j.agrformet.2019.107732, 2019.

Schreiner-McGraw, A. P., Ajami, H., Anderson, R. G., and Wang, D.: Integrating partitioned evapotranspiration data into hydrologic models: Vegetation parameterization and uncertainty quantification of simulated plant water use, Hydrological Processes, 36, https://doi.org/10.1002/hyp.14580, 2022.

Stoy, P. C., El-Madany, T. S., Fisher, J. B., Gentine, P., Gerken, T., Good, S. P., Klosterhalfen, A., Liu, S., Miralles, D. G., Perez-Priego, O., Rigden, A. J., Skaggs, T. H., Wohlfahrt, G., Anderson, R. G., Coenders-Gerrits, A. M. J., Jung, M., Maes, W. H., Mammarella, I., Mauder, M., Migliavacca, M., Nelson, J. A., Poyatos, R., Reichstein, M., Scott, R. L., and Wolf, S.: Reviews and syntheses: Turning the challenges of partitioning ecosystem evaporation and transpiration into opportunities, Biogeosciences, 16, 3747–3775, https://doi.org/10.5194/bg-16-3747-2019, 2019.

Thomas, C. K., Martin, J., Goeckede, M., Siqueira, M., Foken, T., Law, B., Loescher, H., and Katul, G. G.: Estimating daytime subcanopy respiration from conditional sampling methods applied to multi-scalar high frequency turbulence time series, Agricultural and Forest Meteorology, 148, 1210 – 1229, https://doi.org/https://doi.org/10.1016/j.agrformet.2008.03.002, 2008.

Wagle, P., Skaggs, T. H., Gowda, P. H., Northup, B. K., Neel, J. P. S., and Anderson, R. G.: Evaluation of Water Use Efficiency Algorithms for Flux Variance Similarity-Based Evapotranspiration Partitioning in C3 and C4 Grain Crops, Water Resources Research, 57, e2020WR028 866, https://doi.org/https://doi.org/10.1029/2020WR028866, 2021.

Wagle, P., Raghav, P., Kumar, M., and Gunter, S. A.: Influence of water use efficiency parameterizations on flux variance similarity-based partitioning of evapotranspiration, Agricultural and Forest Meteorology, 328, 109 254, https://doi.org/10.1016/j.agrformet.2022.109254, 2023.

Wang, K., Bastos, A., Ciais, P., Wang, X., Rodenbeck, C., Gentine, P., Chevallier, F., Humphrey, V. W., Huntingford, C., O'Sullivan, M., Seneviratne, S. I., Sitch, S., and Piao, S.: Regional and seasonal partitioning of water and temperature controls on global land carbon uptake variability, Nature Communications, 13, https://doi.org/10.1038/s41467-022-31175-w, 2022.

Wang, L., Good, S. P., and Caylor, K. K.: Global synthesis of vegetation control on evapotranspiration partitioning, Geophysical Research Letters, 41, 6753–6757, https://doi.org/10.1002/2014gl061439, 2014.

Wei, Z., Yoshimura, K., Wang, L., Miralles, D. G., Jasechko, S., and Lee, X.: Revisiting the contribution of transpiration to global terrestrial evapotranspiration, Geophysical Research Letters, 44, 2792–2801, https://doi.org/10.1002/2016gl072235, 2017.

Wohlfahrt, G. and Gu, L.: The many meanings of gross photosynthesis and their implication for photosynthesis research from leaf to globe, Plant, Cell and Environment, 38, 2500–2507, https://doi.org/10.1111/pce.12569, 2015.

Wohlfahrt, G., Brilli, F., HÖRtnagl, L., Xu, X., Bingemer, H., Hansel, A., and Loreto, F.: Carbonyl sulfide (COS) as a tracer for canopy photosynthesis, transpiration and stomatal conductance: potential and limitations, Plant, Cell & Environment, 35, 657–667, https://doi.org/10.1111/j.1365-3040.2011.02451.x, 2011.

Zahn, E.: einaraz/PartitioningMethods: Processing Eddy-Covariance Data: Five Evapotranspiration Flux Partitioning Methods (v1.0.1) [Software], https://doi.org/10.5281/zenodo.11510363, https://doi.org/10.5281/zenodo.11510363, zenodo, 2024.

Zahn, E. and Bou-Zeid, E.: Partitioning of water and CO_2 fluxes at NEON sites into soil and plant components: a five-year dataset for spatial and temporal analysis [dataset], https://doi.org/10.5281/zenodo.12191876, 2024.

Zahn, E., Chor, T. L., and Dias, N. L.: A Simple Methodology for Quality Control of Micrometeorological Datasets, American Journal of Environmental Engineering, 6, 135–142, doi:10.5923/s.ajee.201601.20, 2016.

Zahn, E., Bou-Zeid, E., Good, S. P., Katul, G. G., Thomas, C. K., Ghannam, K., Smith, J. A., Chamecki, M., Dias, N. L., Fuentes, J. D., Alfieri, J. G., Kwon, H., Caylor, K. K., Gao, Z., Soderberg, K., Bambach, N. E., Hipps, L. E., Prueger, J. H., and Kustas, W. P.: Direct partitioning of eddy-covariance water and carbon dioxide fluxes into ground and plant components, Agricultural and Forest Meteorology, 315, 108 790, https://doi.org/https://doi.org/10.1016/j.agrformet.2021.108790, 2022.

Zahn, E., Ghannam, K., Chamecki, M., Moene, A. F., Kustas, W. P., Good, S., and Bou-Zeid, E.: Numerical Investigation of Observational Flux Partitioning Methods for Water Vapor and Carbon Dioxide, Journal of Geophysical Research: Biogeosciences, 129, e2024JG008 025, https://doi.org/https://doi.org/10.1029/2024JG008025, e2024JG008025 2024JG008025, 2024.

Zhou, S., Yu, B., Zhang, Y., Huang, Y., and Wang, G.: Partitioning evapotranspiration based on the concept of underlying water use efficiency, Water Resources Research, 52, 1160–1175, https://doi.org/10.1002/2015WR017766, 2016.