https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

Open Access Earth System Science Data Discussions

A continental benchmark dataset for evaluating ecosystem gradient-flux approaches across 47 NEON flux towers

Sparkle L. Malone[1,2], Jaclyn Hatala Matthes[3], Cove S. Sturtevant[4], Angel Chen[1], Roisin Commane[5], Kyle B. Delwiche[6], Ankur R. Desai[7], Christopher R. Florian[4], Jonathan D. Gewirtzman[1], Samuel A. Jurado[1], David E. Reed[1], Jinshu Chi [8], Hiroki Iwata[9], Erik J Lundin[10], Ivan Mammarella[11], Matthias Peichl[12], Masahito Ueyama[13], Camilo Rey-Sánchez[14]

[1]Yale School of the Environment, Yale University, New Haven, 06511, USA
[2] Yale Center for Natural Carbon Capture, Yale University, New Haven, 06511, USA
[3] Harvard Forest, Harvard University, Petersham, 01366, USA
[4] National Ecological Observatory Network, Battelle, Boulder, 80301, USA
[5] Department of Earth & Environmental Sciences, Columbia University, New York, 10027, USA
[6] Department of Environmental Science, Policy and Management, University of California, Berkeley, 94705, USA
[7] Department of Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, Madison, 53706, USA
[8] Earth, Ocean and Atmospheric Sciences Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, 511453, China
[9] Department of Environmental Science, Shinshu University, Nagano, 390-8621, Japan
[10]Abisko Scientific Research Station, Swedish Polar Research Secretariat, Veteskapensväg, 98107, Abisko
[11] Institute of Atmospheric and Earth System Research/Physics, University of Helsinki, Helsinki, 00014, Finland.
[12] Department of Forest Ecology and Management, Swedish University of Agricultural Sciences, Umeå, 90183, Sweden
[13] Graduate School of Agriculture, Osaka Metropolitan University, Osaka, 599-8531, Japan
[14] Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, Raleigh, 27695, USA

Correspondence to: Sparkle L. Malone (sparkle.malone@yale.edu)

Abstract. Benchmark datasets for evaluating profile-based ecosystem flux methods across environmental gradients are currently lacking. This limits method evaluation, cross-site intercomparison, and the expansion of tower-based monitoring for gases not routinely measured by the eddy covariance (EC) method. Here, we present a benchmark dataset derived from 47 terrestrial towers in the National Ecological Observatory Network (NEON), integrating co-located EC, concentration profiles, tower geometry, and canopy structural metrics. We developed a dataset to evaluate the performance of three widely used gradient flux approaches—the modified Bowen ratio (MBR), aerodynamic (AE), and wind-profile (WP) methods—against co-located EC measurements of CO_2 and H_2O. We evaluate how canopy structure, sensor height configuration, and data filtering influence agreement with EC across ecosystems with canopy heights spanning 0.15 to 53 m. Performance varied strongly among approaches and measurement configurations. Across all ecosystems, 11% of height-pair combinations for CO_2 and 19% for H_2O achieved moderate-to-strong concordance with EC, with the MBR approach providing the most consistent performance. Reliable estimates most often arose when both sampling heights were above the canopy (AA) or when one level was above and the other within sparse canopies (AW). Concordance declined with increasing canopy complexity, highlighting persistent challenges in tall and complex forests. Since filtering substantially reduced data availability, we developed an ensemble framework that combined reliable estimates across the three GF approaches and height pairs. Ensemble GF fluxes

reproduced seasonal diel dynamics measured by EC across ecosystems (R^2 = 0.58-0.99), capturing both the seasonal magnitude and direction of net ecosystem exchange. Benchmark analyses show that the GF method can provide robust ecosystem-scale flux information when deployed under favorable structural and micrometeorological conditions. The released benchmark dataset provides a resource for testing new flux algorithms, optimizing sensor placement, benchmarking profile methods, and extending multi-gas tower observations to be more inclusive of gases that are difficult to measure with fast-response EC instrumentation, including CH_4, N_2O, volatile organic compounds, and reactive pollutants.

1 Introduction

We present the first continental benchmark dataset designed to evaluate ecosystem gradient-flux (GF) methods across diverse canopy structures and climates. By integrating standardized observations from 47 NEON towers, the dataset establishes a reusable reference resource for profile-based flux research and future multi-gas observing systems. Such a resource is timely because the exchange of gases between terrestrial ecosystems and the atmosphere regulates climate, hydrology, and air quality (Smith et al., 2013). Measurements of carbon dioxide (CO_2), water vapor (H_2O), methane (CH_4), and other reactive trace gases are therefore foundational for understanding biosphere–atmosphere feedbacks, evaluating ecosystem responses to disturbance and climate change, and constraining Earth system models. Over the past three decades, tower-based micrometeorology has transformed this field by enabling continuous ecosystem-scale flux observations across broad environmental gradients (Baldocchi, 2014, 2003; Hicks and Baldocchi, 2020).

Among available techniques, the eddy covariance (EC) method has become the standard for measuring ecosystem gas exchange (Baldocchi, 2014, 2003). By directly quantifying the covariance between turbulent vertical wind fluctuations and scalar concentrations, EC provides continuous estimates of surface–atmosphere exchange at spatial scales relevant to heterogeneous landscapes. Modern micrometeorology towers measure greenhouse gases (CO_2, CH_4, N_2O), volatile organic compounds (isoprene, monoterpenes), gaseous pollutants (O_3, NO_2, NO), and energy fluxes (sensible heat and latent heat) across diverse biomes from Arctic tundra to tropical forests and wetlands. Although the rise of coordinated networks (e.g., AmeriFlux, FLUXNET, and National Ecological Observatory Network) has made EC central to modern carbon-cycle science, limitations in funding, fast response instrumentation, and access have led to a global network with varying degrees of representativeness across different trace gases (Guenther et al., 2012; Karl et al., 2003; Malone et al., 2022; Reed et al., 2025; Spirig et al., 2005).

As scientific and policy priorities increasingly demand multi-gas monitoring, new strategies are needed to extend the capabilities of existing tower infrastructure. The gradient-flux (GF) method offers one such strategy. Early micrometeorological work employed the GF method, in which fluxes are assumed proportional to a gradient and an eddy diffusivity (K; i.e., the turbulent ability of the air to transport the scalar) (Meyers and Baldocchi, 2015). In the mid-to-late

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

twentieth century, the GF method dominated ecosystem–atmosphere flux studies by measuring the difference in a scalar at two heights above the canopy and estimating K from profiles or similarity relations (Cellier and Brunet, 1992; Meyers and Baldocchi, 2015; Simpson et al., 1998). Numerous studies have illuminated the essential assumptions and limitations of the GF method in ecosystems with complex canopies (Raupach, 1979; Raupach and Legg, 1984; Stewart and Thom, 1973). Yet, the constraints imposed by the physical infrastructure required to apply this method (Wagner-Riddle et al., 2005) limited its reliable use.

In principle, the GF method could greatly expand atmospheric observing capacity by enabling flux measurements for gases that are difficult to measure with EC. In practice, however, the GF method is sensitive to assumptions that are often violated. Their performance depends on accurate estimation of K, sufficient gradient strength relative to instrument noise, atmospheric stationarity, limited advection, and measurement heights that adequately sample the relevant source area. Tall or structurally complex canopies further complicate application because turbulence within the roughness sublayer departs from classical similarity theory (Zhao et al., 2019), while vertical distributions of canopy and soil sources can generate counter-gradient behavior or footprint mismatch among sampling levels. These constraints have limited widespread adoption and left a basic operational question unresolved: when and where does the GF method produce reliable ecosystem-scale fluxes?

Today, the widespread availability of EC provides a robust benchmark for assessing GF performance. It is now possible to define the operational limits of the GF method and strategically deploy it to extend our understanding of trace gas fluxes beyond the constraints imposed by EC alone. The National Ecological Observatory Network (NEON), funded by the National Science Foundation, has uniform micrometeorological instrumentation at 47 terrestrial sites across the United States. The NEON tower sites represent a wide diversity of ecoclimate regions and canopy structures (Kao et al., 2012; LaRue et al., 2022; Metzger et al., 2019). Each site has an EC at the top of the tower for CO_2 and H_2O, and a concentration profile extending down into the canopy, providing multiple sampling height pairs for GF calculation. Micrometeorology, paired with canopy structure and measurements at multiple levels above and within the canopy, provides the infrastructure to evaluate the extent to which the GF method can reliably produce ecosystem fluxes.

In this study, we utilize co-located EC and GF instrumentation (Bolinius et al., 2016; Wu et al., 2015; Zhao et al., 2019), along with measures of canopy complexity (e.g., canopy lidar and spectral information). Specifically, we compare different methodological approaches to calculate GF fluxes, including the modified Bowen ratio (MBR), the aerodynamic (AE), and the wind profile (WP) approaches (Bolinius et al., 2016; Mayer et al., 2011; Meredith et al., 2014; Meyers et al., 1996). While all three GF approaches derive fluxes from mean concentration gradients, they differ in how they determine K. The MBR assumes that K is identical for a tracer and target gas and that vertical concentration gradients are consistent with the turbulent flux (Meyers et al., 1996). This assumption allows calculation of a target flux when only the tracer flux is measured with EC. Typically, the tracer gas is easier to measure (e.g., H_2O and CO_2) than the target gas (e.g., CH_4, N_2O, isoprene, monoterpenes, O_3, NO_2, NO). With the AE, K is derived from wind and turbulence statistics. This approach assumes steady-state, horizontally homogeneous conditions and is sensitive to atmospheric stability and surface roughness parameters (Dyer and Hicks, 1970; Monteith and Unsworth, 2013). The WP approach is a simplification of the AE that derives K directly from the measured wind speed profile and surface roughness, without assuming external turbulence parameters. The WP assumes steady-state, horizontally uniform conditions, and errors in estimating displacement height or roughness propagate into flux estimates. The WP is less accurate in complex terrain or under low turbulence (Stull, 1988).

We produced a standardized, multi-site gradient-flux benchmark dataset to evaluate ecosystem-atmosphere exchange methods across NEON towers. The uncertainty and reliability of flux estimates from three GF methods (MBR, AE, and WP) were evaluated using CO_2 and H_2O EC fluxes measured at the tops of NEON towers as benchmarks. We assess how canopy structure (e.g., height, gap fraction, top-of-canopy roughness, and heterogeneity) influences the likelihood that a given sampling height pair will produce reliable ecosystem fluxes. Ultimately, our goal is to determine where and when the GF method can be applied at NEON micromet towers, or at towers with similar characteristics in sampling time and measurement height. With the benchmark dataset, we evaluate three guiding questions:

(Q1) How well does the GF method perform across different ecosystems?

(Q2) Which sampling height pairs, in relation to canopy properties, provide reliable ecosystem fluxes?

(Q3) Does the GF method provide a valid ecosystem assessment?

Although much of what we know about the limits of the GF method is rooted in the foundational history of the EC method (Raupach, 1979; Raupach and Legg, 1984; Stewart and Thom, 1973), it is now possible to record data from more instruments over extended periods and with greater precision than previously available. The NEON data provide an opportunity to review and refine our understanding of the GF method, enabling us to see how it can be used to understand ecosystem exchange rates. Our benchmark dataset enables reproducible evaluation of profile-based ecosystem flux methods across continental gradients in canopy structure, climate, and tower configuration.

2 Methods

2.1 Study Sites

NEON's standardized observations and data-processing suite is explicitly designed for inter-site comparability and for analysing feedback across multiple spatial and temporal scales. Flux towers have multiple sampling levels, each outfitted with a suite of sensors mounted on perpendicular boom arms. The number of sampling levels on the tower varies from 4 to 8, depending on the site's mean canopy height. Towers were designed to be taller than the vegetation canopy at each site. Tower sampling heights were categorized by the measurement heights of the two levels used to calculate the GF (Table 1). Measurements were inclusive of both sampling levels above the canopy (AA) or one measurement height within the canopy (AW). Although AW is not considered ideal, when canopies are sparse and allow sufficient connectivity, fluxes have been shown to provide reliable assessments of ecosystems (Meredith et al., 2014). We also included an additional modifier indicating whether a measurement height was above and most adjacent (+) or below and most adjacent (-) to the mean canopy height (Figure 1B). Across 47 towers, 208 measurement pairs were AA, and 237 pairs were AW. Tower sensors continuously collected data to capture patterns and cycles over periods ranging from seconds to years. The NEON implements a

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

Open Access Earth System Science Data Discussions

comprehensive quality management system, including frequent sensor calibration and data quality checks to minimize measurement errors and ensure high-quality data collection (Sturtevant et al., 2022).

Table 1: The gradient flux (GF) measurements were categorized into eight groups relative to the mean canopy height, with a modifier indicating whether a sampling height was directly above and most adjacent (+) or below and most adjacent (-) to the mean canopy height.

Canopy Level	Description	Sampling Height Pairs	Towers
AA+	Both measurements above the canopy, with one of them directly above the canopy	114	47
AA	Both measurements above the canopy	94	43
AW+-	One measurement directly above and one directly below the canopy height	32	32
AW+	One measurement directly above and one within the canopy	46	27
AW-	One measurement above and one directly below the canopy height	69	32
AW	One measurement above and one within the canopy	90	27

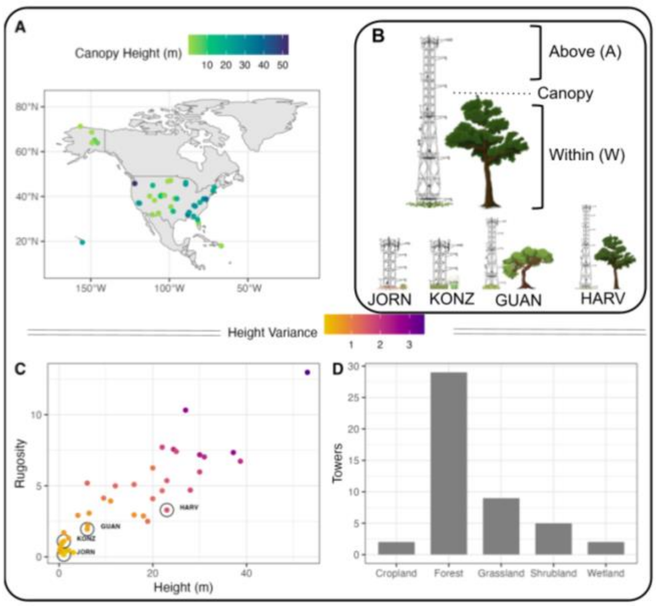


Figure 1: NEON tower sites (A) are distributed across a range of canopy heights with measurement levels above and below the canopy (B). Across sites, canopy complexity increases with height (C). Sites are also distributed across five ecosystem types (D).

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

The NEON sites encompass 47 ecosystems, exhibiting wide variation in structural properties that influence the utility of the GF approach (Figure 1). Forests account for 62% (29 towers) of the towers and include tropical, temperate, and boreal forests, as well as savannas. Canopy complexity across forest types typically suggests that more diverse canopy structures indicate a more open canopy compared to heterogeneous tall canopies (Wang et al., 2024). Short-stature ecosystems include shrublands (5 towers) and grasslands (9 towers). While most towers represent upland ecosystems, temperate wetlands are also represented (2 towers). Within this wide variety of ecosystem types, canopy heights range from 0.15 to 53 meters. While GF methods were evaluated across all sites, we further selected a subset of the 47 tower sites, distributed across canopy variability, to examine site-level effects. The sites include the arid grassland in the Jornada Basin (JORN), the tallgrass prairie in Konza Prairie (KONZ), the tropical evergreen forest at Guanica Forest (GUAN), and the temperate mixed forest at Harvard Forest (HARV).

2.2 Instrumentation

Fast response fluxes of CO_2 (μmol m^{-2} s^{-1}), latent heat flux (W m^{-2}), sensible heat flux (W m^{-2}), and friction velocity, u* (m s^{-1}) were measured using a gas analyzer (LI-COR - LI7200) and sonic anemometer (Campbell Scientific CSAT-3) at 20 Hz at the top of the towers. Eddy covariance processing at a 30-minute timescale was done following NEON protocols (Metzger et al., 2017) with de-spiking of data (Brock, 1986; Starkenburg et al., 2016), planar-fit coordinate rotation (Yuan et al., 2011), and corrections for the infrared gas analyzer (Welles and McDermitt, 2005), time lag (Rebmann et al., 2012), sonic temperature (Schotanus et al., 1983), and high-frequency spectral loss (Metzger et al., 2016).

Meteorological data included aspirated air temperature (Thermometrics Climate RTD 100 Ω Probe) and horizontal wind speed and direction (Gill Wind Observer II) at multiple profile heights at each site. Atmospheric pressure P (kPa; Vaisala PTB 330) and relative humidity (Vaisala HMP 155) were measured at the mid-tower and the tower top, respectively. Dry air mixing ratios of CO_2 (μmol mol^{-1}) and H_2O (mmol mol^{-1}) were measured at multiple profile heights at each site using a Picarro G2131-i NEON-specific analyzer with additional upgrades to the CH_4 channel. The gases, CO_2 and H_2O, were measured at a nominal rate of 1 Hz. The means and variances of calibrated mixing ratios were calculated at each sampling level (6-minute sampling periods). Air was sampled from inlets at each tower sampling level, starting at the top and proceeding to the next lower inlet until reaching the bottom, then returning to the top.

The NEON towers use high-accuracy sequential sampling over the tower profile. The mixing ratio pairs used for gradient differences spanned at least 20 minutes and had a 10-minute dwell time per sampling level (4-min flux, 6-min sampling). Although sequential sampling is less than ideal, studies have shown that only minor errors are introduced into flux measurements when a single analyzer measures non-simultaneously at alternating heights (Griffith et al., 2002; Kamp et al., 2020; Nelson et al., 2019; Tanimoto et al., 2014). Meteorological data were available as 1- or 2-minute averages, depending on the variable, and the shortest available duration was used. Meteorological variables were aligned with the paired concentration measurement window by averaging them over the 20-minute sampling window. This approach, using linear

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

regression for large time delays, has been used in other studies (Hall et al., 1999). Turbulent flux measurements reported as 30-minute averages were filtered using NEON's final quality flag before linearly interpolating to the time midpoint of the combined sampling window's observed paired-level gradient mixing ratios. Additionally, due to interpolation, GFs could be calculated only when both the preceding and succeeding half-hourly eddy covariance fluxes were available. All data were downloaded using the *neonUtilities* package in R (Table 2).

Table 2: We used data from the National Ecological Observatory Network (NEON). This continental-scale research platform collects long-term, open-access data on animals, soil, water, and the atmosphere across the U.S. to understand how ecosystems are changing.

Data Product	NEON Product ID	Reference
Relative Humidity	DP1.00098.001	(NEON, 2025g)
2D wind speed and direction	DP1.00001.001	(NEON, 2025a)
Photosynthetically active radiation (PAR)	DP1.00024.001	(NEON, 2025f)
Bundled Eddy Covariance	DP4.00200.001	(NEON, 2025b)
Vegetation Indices	DP3.30026.001	(NEON, 2025h)
Leaf area index (LAI) mosaics	DP3.30012.001	(NEON, 2025e)
Ecosystem Structure	DP3.30015.001	(NEON, 2025c)
Elevation-LiDAR	DP3.30024.001	(NEON, 2025d)
Discrete Return LiDAR point cloud	DP1.30003.001	(NEON, 2025d; Wang et al., 2024)

2.3 The Eddy Covariance Method

With the EC method, the net transport between the surface and the atmosphere is assumed to be one-dimensional, and the covariance between turbulent fluctuations of the vertical wind and the quantity of interest is used to calculate the vertical flux density. Measurements were typically made in the atmospheric surface layer, where normalized fluxes are constant with height, primarily represent fluxes from the underlying surface, and shear-driven turbulence is the dominant transport mechanism. The EC method involves fast-time-resolution measurements of turbulence and trace gases, with measurement scales typically ranging from 1 to 3 km², depending on measurement height, vegetation canopy height, and turbulence characteristics. Low-turbulence conditions were determined following Foken and Leclerc (2004), assigning quality flags and filtering when advective transport terms dominate and lead to underestimates of surface fluxes from turbulent measurements (Foken and Leclerc, 2004). Turbulent fluxes were expressed with time average components represented with overbars and deviations from the mean due to turbulence represented with prime notation. Net ecosystem exchange includes the turbulent flux, advection, and storage. In our study, advection was considered negligible, and storage was not included. Although we may be biasing the ecosystem flux by not including storage, both the flux gradient and the eddy covariance represent turbulent flux; as such, they are both missing the storage component and thus are comparable.

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

2.4 Flux Gradient Method

The GF method utilized the difference in concentrations measured at two heights and the assumption that turbulent transfer of scalars is analogous to molecular diffusion. The standard approach is expressed as:

$$F = -\rho_d \, K \, \frac{\partial C}{\partial z} \tag{1}$$

where F is the target turbulent flux (e.g., CO_2, H_2O, or CH_4; µmol m^{-2} s^{-1}), ρ_d is the dry air density (mol m^{-3}), C is the scalar dry air mixing ratio (µmol mol^{-1}) at height z (m), and K is eddy diffusivity (m^{2} s^{-1}). Using dry air mixing ratios eliminates density effects. One challenge of the gradient-flux method is to parameterize the eddy diffusivity (K), which can be determined from the Monin-Obukhov similarity theory (Dyer and Hicks, 1970) or, as suggested by Miyata et al. (2000), from the friction velocity derived from a sonic anemometer (Miyata et al., 2000).

2.4.1 Modified Bowen Ratio

Fluxes were estimated following the MBR approach (Eq. 2) by expressing Eq. 1 for two fluxes, F_1 and F_2, taking the ratio, leading K to cancel out, and rearranging to:

$$F_2 = F_1 \, \frac{(\Delta C_2)}{(\Delta C_1)} \tag{2}$$

Where F_2 is the target flux, F_1 is a known flux, ΔC_2 is the target gas dry air mixing ratio gradient difference, and ΔC_1 is the known gas dry air mixing ratio gradient difference measured across the same height profile. The assumption underlying this approach is that the eddy diffusivity is identical for both the tracer (F_1) and the target gas (F_2). If F_1 is known from EC measurements, F_2 can be determined (Moncrieff et al., 1997). Using the MBR method removes the need to invoke Monin-Obukhov similarity for different stabilities, keeping the physics simple. Previous studies have used these methods to calculate ecosystem-scale fluxes for trace gases (e.g., methane (Werner et al., 2003; Zhao et al., 2019), hydrogen (Meredith et al., 2014), carbonyl sulfide (Commane et al., 2015), and mercury (Obrist et al., 2021).

2.4.2 The Aerodynamic Method

In the aerodynamic model, K is given as:

$$K = \frac{k \, u^{*}(z-d)}{\phi_h(\zeta)} \tag{3}$$

where K is assumed to be the same for all variables (H_2O and CO_2), the von Karman constant (k) is 0.4 (Stull, 2012), $u*$ is the friction velocity and is measured by the fast response turbulent flux data, z is the mean of measurement height calculated as the geometric mean between the two levels, and d is the displacement height. Canopy height (h) was calculated aerodynamically from the turbulence statistics (Pennypacker and Baldocchi, 2026), and displacement height was estimated as 0.66h. The ϕ_h (ζ) is the stability correction function, and ζ is the Obukhov stability parameter calculated as the ratio of

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

effective height (z-d) and Obukhov length (L). The φ_h is assumed to be equal to the stability correction function for CO_2 and H_2O, and calculated as (Lee et al., 2018):

for -5 < ζ < 0:

$$\varphi_h = (1 - 16\zeta)^{-1/2} \tag{4}$$

for 0 <= ζ < 1:

$$\varphi_h = 1 + 5\zeta \tag{5}$$

The Obukhov length is given as:

$$L = \frac{-u_*{}^3 T}{kg(\overline{w'\theta'_v})} \tag{6}$$

where T is the surface air temperature measurement taken from the lowest tower height, g is gravitational acceleration (9.8 m s^{-2}), and θ_v is the virtual potential temperature.

2.4.3 The Wind Profile Method

The eddy diffusivity for the wind profile method (K_{wp}) was calculated at each tower level as:

$$K_{wp} = \frac{k^2\,(z-d)\,\underline{u}}{\varphi_h\, ln\frac{z-d}{z_o}} \tag{7}$$

where k is the von Karman constant, z-d is the effective height as calculated in the previous section, $\underline{u}$ is the average wind speed, and z_o is the roughness length estimated as $0.1h$.

2.5 Filtering Gradient Fluxes

Gradient fluxes were filtered using a signal-to-noise ratio (SNR) based on the difference in concentration (ΔC) of the target gas (Eq. 8):

$$SNR = \frac{|\Delta C|}{sd(\Delta C)} \tag{8}$$

We randomly sampled concentration measurements within the sampling period to estimate the standard deviation in $\triangle$C. A conservative threshold of 3 was used at every site, based on the stabilization of the error between EC and GF for SNR values above this threshold. Depending on the technique, an SNR for both the target and tracer gases (TSNR) was required (MBR). In addition to filtering for SNR, we also filtered GFs using the u* threshold defined for each site by AmeriFlux (Gu et al., 2005) at the top of the towers. The u* threshold indicated when shear-driven turbulent mixing was insufficient (Gu et al., 2005). Filtering GFs resulted in substantial data loss across approaches, ranging from 70% to 90%, for both CO_2 and H_2O. The SNR was the strongest filter, and u* was typically associated with less than 50% data loss. The TSNR, used for the MBR, filtered an additional ~20-30% of the remaining data on average. As a result of the TSNR, more data was filtered for the MBR

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

than for the AE and WP. Filtered data also varied by sampling height pairs. Fewer data were filtered when at least one measurement was above the canopy (AA or AW).

2.6 Evaluation of Gradient Fluxes

We compared filtered GFs to the EC method. We determined which sampling height pairs yielded reliable data using Lin's concordance correlation coefficient (CCC), a statistic that assesses the agreement between two continuous measurement sets (Lin, 1989). The CCC is often used to evaluate the reliability against a benchmark. A CCC value close to 1 indicates high positive agreement, meaning a strong linear relationship with slight bias. A CCC value close to -1 suggests strong negative agreement, meaning a strong linear relationship, but the two datasets are mirror images of each other. A CCC value close to 0 implies a very weak linear relationship. To test for significant differences in CCC values, we used the Kruskal-Wallis rank sum test in R (Okoye and Hosseini, 2024; Ostertagova et al., 2014). This nonparametric statistical test was used to determine if there were statistically significant differences between the medians of GF approaches or sampling height pairs. The Kruskal-Wallis test ranks all the data across all groups, then sums the ranks within each group and compares the sums to determine whether they differ significantly.

To assess an appropriate concordance threshold for identifying reliable sampling-height pairs (RSHP), we conducted a sensitivity analysis using three candidate cutoffs for the agreement between gradient fluxes and eddy covariance (EC), CCC ≥ 0.3, 0.5, and 0.7 (Table A1). For each cutoff, we determined the percentage of sampling height pairs retained, the level of sign agreement, the mean and absolute flux biases, the correlation, and the root-mean-square error (RMSE) between GFs and EC. The lower threshold (CCC >= 0.3) was a permissive screen emphasizing directional consistency, the upper threshold (CCC >= 0.7) was a conservative screen favoring low bias, and the intermediate threshold (CCC >= 0.5) was a candidate balance point. We then selected the threshold that provided the strongest trade-off between retaining enough sampling-height pairs for analysis and achieving sufficiently strong flux concordance with EC. Based on this comparison, CCC >= 0.5 was retained as the primary operational threshold for moderate agreement (Landis and Koch, 1977) between EC and GF, balancing concordance with data loss. After calculating the CCC based on EC and filtering the CCC values below the threshold, we assess the amount of information retained in the remaining data.

2.7 Canopy Influence on Reliable Sampling Height Pairs

We explored how canopy characteristics (Table 3) affect GF reliability across different sampling height pairs. Based on the CCC, we identified reliable/non-reliable sampling height pairs at each tower site. Next, we gathered canopy-structure information for each site from the NEON Airborne Observation Platform (AOP) data and the 30 m NEON structural diversity dataset, which were derived from the NEON AOP Discrete Return LiDAR point cloud product (Wang et al., 2024). A machine

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

learning approach was employed to assess the impact of canopy structure and function on the reliability of sampling height pairs. First, data were divided into training (65%) and testing (35%) sets using stratified random sampling to maintain the distribution across all variables. A preliminary step included variable selection (Table 3). We fit random forest (RF) models using a series of candidate independent variable subsets, employing a modified backward elimination procedure with the VSURF package in R (Genuer et al., 2015). The fit of each model was evaluated using the out-of-bag mean square error (OOB MSE), and variable importance was computed as the increase in prediction error resulting from permuting a particular predictor. While random forest models are robust to multicollinearity (Genuer et al., 2015), we limited the inclusion of highly correlated variables to ensure that each variable in the final model improved model fit and did not affect other variables. Following variable selection, we fit a final model using the randomForest package (Liaw and Wiener, 2002; R Core Team, 2021). Each tree was built from a balanced bootstrap sample: 305 observations from sampling height pairs where CCC < 0.5, and 305 from sampling height pairs where CCC > 0.5, rather than sampling in proportion to class imbalances in the data. Model fit was evaluated using the average of the 500 out-of-bag (OOB) mean squared errors (MSEs) from the final model, and variable importance was calculated as the average rank of each predictor variable across the 500 models. We explored the confusion matrix and the no-information rate, which represents the accuracy you would achieve by always predicting the most frequent class, without using a model. We compared model accuracy to the no information rate to evaluate model performance.

Table 3: Explored drivers of reliable sampling height pairs across NEON towers, including canopy attributes (Wang et al., 2024) that quantify measures of complexity and measurement height relative to the mean canopy height.

Category	Metrics	Description	Unit
Height	Mean Canopy Height (CHM)	Mean of maximum height in a 1 m^2 grid	m
Density	Vegetation Area Index (VAI)	Vegetation area per unit of ground area	m^2 m^{-2}
Openness	Deep Gap Fraction (DGF)	Fraction of 1m^2 canopy gaps in the plot	1
Exterior Complexity	Top Rugosity (TopRugosity)	Standard deviation of outer canopy heights in 1 m^2 of plot	m
	Rumple Index (RumpleIndex)	Ratio between the canopy surface area and its projected area on the ground	m^2 m^{-2}
Interior Complexity	Gini Coefficient Index (GiniCoeff)	The Gini coefficient of point height	
	Foliage Height Diversity (FHD)	The distribution of canopy cover among vertical canopy layers is quantified using a diversity index.	
	Standard Deviation of Height (SDH)	Standard deviation of point height	m

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

	Mean of Standard Deviation of Height (MSDH)	Mean of the standard deviation of the point height in 3 × 3 m grid	m
	Standard Deviation of Standard Deviation of Height (SDSDH)	Standard deviation of the standard deviation of height in 3 × 3 m grid	m
Ancillary	Area	Area of the bounding box of points in the NEON AOP Discrete Return LiDAR product	m^2
	Point Density	Number of points per unit of ground area in the NEON AOP Discrete Return LiDAR product	Pts m^{-2}
Canopy Function	The normalized difference vegetation index (NDVI)	Normalized ratio of NIR and Red bands	
	Leaf area index (LAI)	Ratio of upper leaf surface area to ground area	
	Photochemical Reflectance Index (PRI)	Normalized ratio of bands sensitive to xanthophyll pigments	
	Enhanced Vegetation Index (EVI)	Ratio of NIR and Red bands that also includes the Blue channel for better aerosol characterization	
	Soil Adjusted Vegetation Index (SAVI)	normalized ratio of NIR and Red bands with correction factors to minimize soil contribution	
Measurements Distance	Upper distance to the canopy (Relative Dist A)	Relative Distance of the upper sampling height to the canopy	m
	Lower distance to the canopy (Relative Dist A)	Relative Distance of the lower sampling height to the canopy	m

2.8 Gradient Flux Ensemble

We explored different GF ensembles across sampling height pairs and approaches:
Bayesian fusion, inverse RMSE weighted, approach weighted, mean, median, and uncertainty-weighted median. The Bayesian fusion variant used a precision-style weighting scheme proportional to the product of the approach-level and observation-level inverse-variance proxies. The inverse-RMSE weighted ensemble used a weighted mean with weights equal to 1/RMSE. Approach-specific weighting used approach-level mean RMSE, estimated separately by gas and approach across the RSHP-filtered dataset, to assign each method a fixed reliability weight. The uncertainty-weighted median used inverse RMSE as the observation weight. The mean and median used all retained GF estimates equally. When weights were missing or non-positive,

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

the algorithm reverted to the unweighted mean for that half-hour. We evaluated each ensemble against EC reference fluxes at the site-by-gas level. For each site, gas, and ensemble method, we retained only half-hourly comparisons with non-missing EC and ensemble fluxes that met the requirement of more than 20 observations. We then calculated three performance metrics: R^2, CCC, and RMSE between EC and ensemble GFs. Although this is the first study to ensemble GFs across approaches and sampling height pairs, ensemble models have been shown to provide more robust and accurate estimates than single-model approaches (Willcock et al., 2020). Ensembles can also be useful for indicating uncertainty, particularly when validation data are unavailable.

2.9 Ecosystem Assessment

To evaluate whether the ensemble GF can provide ecologically meaningful information, we fit seasonal diel curves for both CO_2 and H_2O to assess fluxes over a 24-hour cycle. Daily changes in sunlight, temperature, and biological activity primarily drive diel curves for CO_2 and H_2O. Curves reveal an ecosystem's metabolic function by showing daily fluctuations in photosynthesis, respiration, and transpiration. Diel curves were fitted on ensemble GF data by season (winter, spring, summer, and autumn) at each site using the loess function in R (Jacoby, 2000). Using the diel curves, we measured the differences in total daily flux between the EC and the ensemble GF. The goal was to compare the pattern and magnitude, which should be similar between the EC and the ensemble GF, to ensure this information is a reliable data stream.

2.10 Methods Summary

This study was designed to explore when and where the GF method reliably estimates ecosystem fluxes of CO_2 and H_2O for the NEON flux towers (Figure 2). Confidence in the GF method was determined by its concordance with the EC method. While we don't expect identical flux measurements between the EC and GF, we assess reliability by comparing the ecological implications of the measurements, which we can quantify effectively using the CCC. To determine the limits of the GF method, we begin by filtering based on u* and SNR. Next, we determined the CCC for each sampling height pair. We created an indicator to identify data from sampling heights with a CCC ≥ 0.5, balancing agreement with the EC while minimizing data loss. With the data that met the CCC threshold, we then explored the distribution of data across GF approaches (MBR, AE, WP), canopy measurement levels (AA and AW), seasons (winter, spring, summer, and autumn), and hours (0-23). We developed a model to explore the impact of canopy attributes on which pairs of sampling heights yielded reliable fluxes. Across ecosystems, where reliable GFs were available from different approaches and sampling height pairs, we ensembled all reliable GFs. We then compared the ensemble GF to the EC method and measured the ensemble's concordance. Next, we performed ecosystem assessments by fitting seasonal diel curves for each site. To assess agreement, we compared diel curves from the EC and ensemble GF methods.

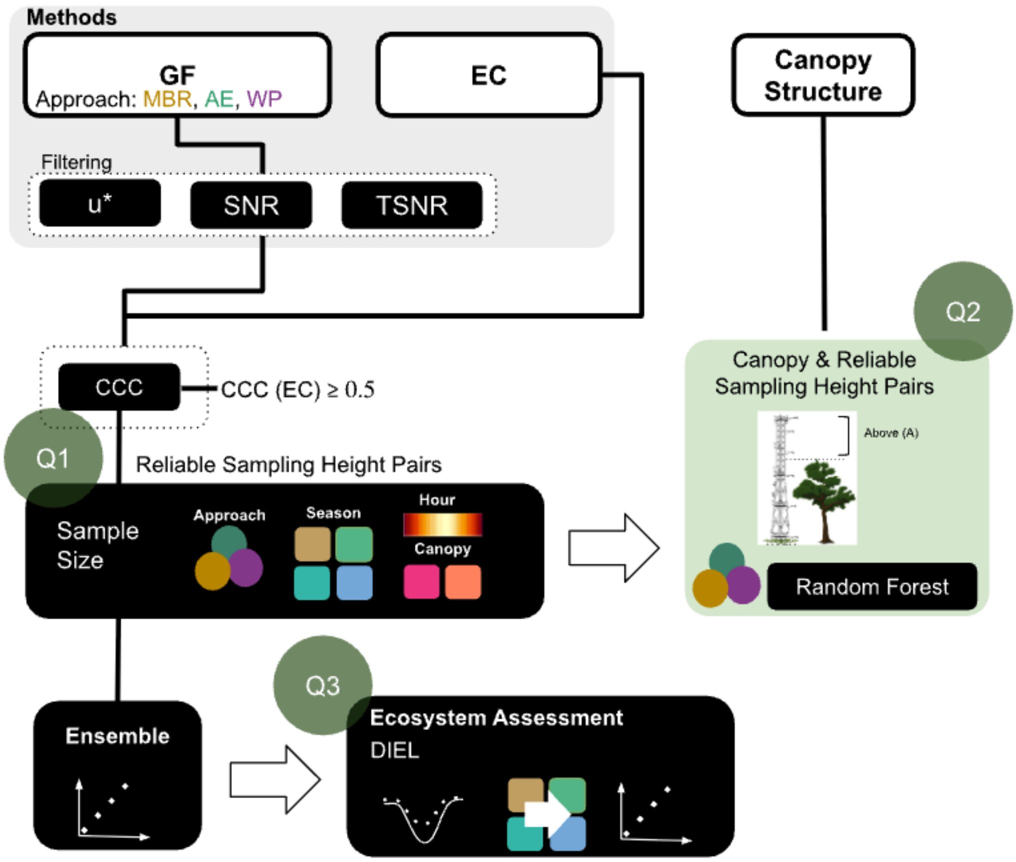


Figure 2: This workflow was designed to explore when and where the gradient flux (GF) method produces reliable ecosystem fluxes of CO_2 and H_2O, using the eddy covariance (EC) method as a benchmark. Guiding questions include: How well does the GF method perform across different ecosystems (Q1)? Which sampling height pairs provide reliable ecosystem fluxes (Q2)? Does the GF method provide a valid ecosystem assessment (Q3)? We used a moderate concordance (CCC) threshold of 0.5 when comparing GFs to EC (Table S1).

3 Results

Most terrestrial NEON tower sites (45 out of 47) had GF data after filtering for u* and SNR. The GFs were distributed across approaches and canopy levels (AA and AW). While all GFs captured combinations of ecosystem components (soil and canopy) to varying degrees, the challenge was to determine which sampling height pairs were representative of the ecosystem-scale flux by measuring their concordance with the EC method.

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

3.1 Gradient Flux Method Evaluation

The CCC for all sampling height pairs ranged from -0.66 to 0.97, with notable differences observed between approaches (MBR, AE, WP) and canopy levels (AA and AW; Figure 3). The range in CCC observed suggests that some sampling height pairs yielded GFs that aligned with the EC method. On average, the mean CCC for CO_2 was higher for the MBR (CCC = 0.27; $p < 0.001$) than the AE (CCC = 0.09; $p < 0.001$) and WP (CCC = 0.13; $p < 0.001$; Figure A1). There was also a significant difference in concordance between canopy levels AA (CCC = 0.20; $p < 0.001$), AW (CCC = 0.13; $p < 0.001$; Figure A1). Similar patterns were observed for H_2O, where there was a significant difference between the CCC of each GF approach and canopy level ($p < 0.001$; Figure A1). The MBR had a higher mean CCC (CCC = 0.29; $p < 0.001$) compared to the AE (CCC =0.11; $p < 0.001$) and WP (CCC = 0.24), and the canopy level AA had a higher mean CCC (CCC = 0.26; $p < 0.001$) compared to AW (CCC = 0.17; $p < 0.001$; Figure 3). Site-based results are shown in Supplement A (Figures A2 and A3).

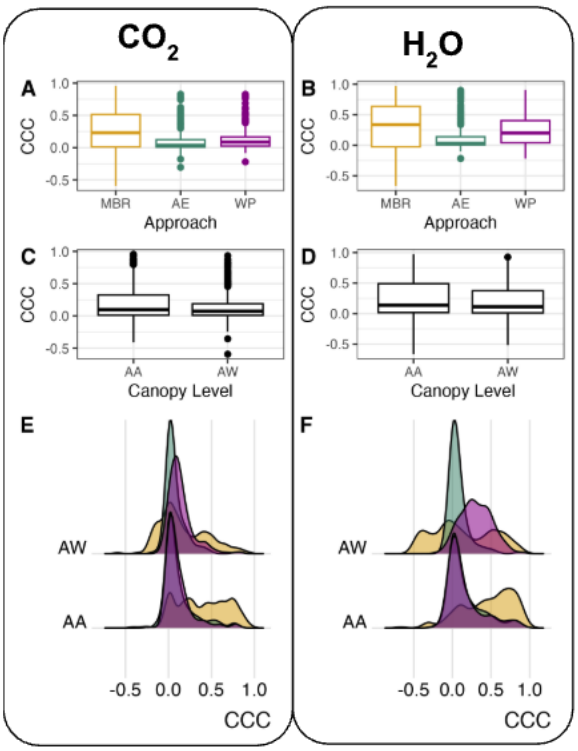


Figure 3: The concordance correlation coefficient (CCC) of the GF compared to EC for sampling height pairs. Concordance varied by approach (MBR, AE, WP: tan, green, and purple, respectively) for CO_2 (A) and H_2O (B) and by canopy level (AA and AW) for CO_2 (C) and H_2O (D). The distribution of CCC values by approach and canopy level for CO_2 (E) and H_2O (F) showed wide variation in concordance across sampling-height pairs.

Of the 47 terrestrial NEON tower sites, 38 sites had reliable sampling height pairs with a CCC ≥ 0.5 for CO_2 (Table A1; Figure 4). Across all sites, sampling height pairs, and approaches, 11% of sampling height pairs produced reliable GFs for CO_2 (Figure A4). The mean number of sampling height pairs at a site with a CCC ≥ 0.5 was 3.8 (2.6 S.D.) for CO_2 and 5.7 (3 S.D.)

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

for H_2O. Most of the reliable sampling height pairs were from the MBR approach, followed by the AE and WP for CO_2. The most reliable sampling height pairs were above the canopy, with just 28% from the AW for CO_2. The CCC for a reliable sampling height pair (CCC ≥ 0.5) was similar for AA (CCC = 0.69; R^2= 0.58) and AW (CCC = 0.67; R^2= 0.53) for CO_2. More sites had reliable sampling height pairs for H_2O, and 30 sites provided reliable GFs for both gases. Across all sites and sampling height pairs, 19% of sampling pairs were reliable for H_2O (Figure A4). Similar to patterns observed for CO_2, the most reliable sampling height pairs were for the MBR, followed by the AE and the WP. The most reliable sampling height pairs for H_2O were AA and AW (Figures S3 and S4). The concordance was greatest for the AA (CCC = 0.72; R^2= 0.59) compared to the AW (CCC = 0.65; R^2= 0.50) across and within NEON tower sites for H_2O.

Across approaches, sites had reliable CO_2 and H_2O data in at least one of four seasons (winter, spring, summer, and autumn; Figure A5). For CO_2, observations were most common in summer (34%; 11% S.D.) and autumn (25%; 7 S.D.), with the spring (24%; 9 S.D.) and winter (17%; 8 S.D.) undersampled. For H_2O, most flux data were from the spring (31%; 12 S.D.), and the summer (24%; 7 S.D.), autumn (21%; 6 S.D.), and winter (23%; 9 S.D.) were undersampled on average. The patterns in percent coverage by hour were consistent across seasons and suggest that early-morning and midday fluxes were undersampled for CO_2, and that early hours were undersampled for H_2O (Figures A5B and A5C). The distribution of GFs across seasons (Figure A5) and hours (Figures A6 and A7) is shown in Supplement A.

Across all ecosystem types, the MBR provided reliable sampling height pairs, and the AE and WP were important for tall canopies. The MBR produced reliable GFs when both measurement levels were above the canopy (AA), and the AE and WP were especially important for reliable cross-canopy fluxes (AW). As a result, the MBR was the most frequently retained approach across ecosystem type, and the AE and WP were essential when cross-canopy fluxes were available, enabling reliable GFs. Still, the size of GF datasets was larger in short-stature ecosystems, where the rate of reliable sampling height pairs above the canopy was higher (AA; Figure A8). On average, the dataset size across the approaches indicates that AE and WP were larger (Figure A9 and Figure A10). Even though the AE and WP had larger datasets and were associated with forests, these ecosystems also had the smallest datasets on average (Figure A8).

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

Open Access Earth System Science Data Discussions

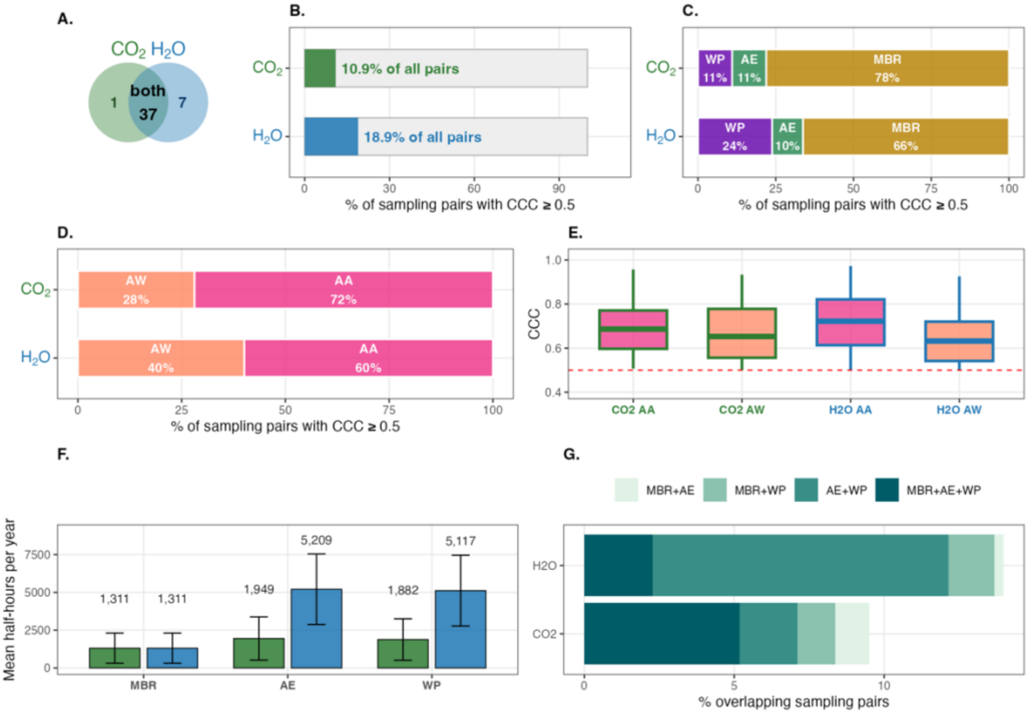


Figure 4: Summary of sampling height pairs across NEON tower sites. The number of sites with sampling height pairs with a CCC ≥ 0.5 for CO_2 and H_2O (A); the percent of sampling height pairs with a CCC ≥ 0.5 (B); by approach (MBR, AE, WP; C); and by canopy level (D). The distribution of sampling height pairs with a CCC ≥ 0.5 by canopy level (E). The mean size of flux datasets by canopy level (F) and the percent overlap of measurements by approach (G).

3.2 Canopy Influence on Reliable Sampling Height Pairs

Structural attributes of ecosystems influenced which pairs of sampling heights had a CCC ≥ 0.5 (Figure 5). With the GF approach, the mean canopy height within the tower footprint, the standard deviation of canopy height (SDSDH), and the distance between the measurement height and the canopy at the upper (Relative Dist A) and lower (Relative Dist B) measurement levels influenced the reliability of sampling height pairs. The random forest model accuracy was 95%, which was significantly different from the no-information rate (0.85; $p < 0.001$). Model precision was balanced for the CCC ≥ 0.5 (precision = 0.95) and the sampling heights with CCC < 0.5 (precision = 0.92). The sensitivity analysis revealed that the probability of a sampling height pair having a CCC ≥ 0.5 was highest for the MBR approach when measurements were above the canopy (i.e., AA; Figure 5). In contrast, AE and WP performed best when the top sampling height pair was above the canopy, and the bottom measurement was within the canopy (i.e., AW). Measurement levels farther from the mean canopy height had a higher probability of producing a reliable sampling height pair, except for the WP approach, where a measurement level within the canopy, closest to the top, also had a higher probability. Sampling height pairs for ecosystems with a lower canopy variance and shorter mean canopy height had a higher probability.

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

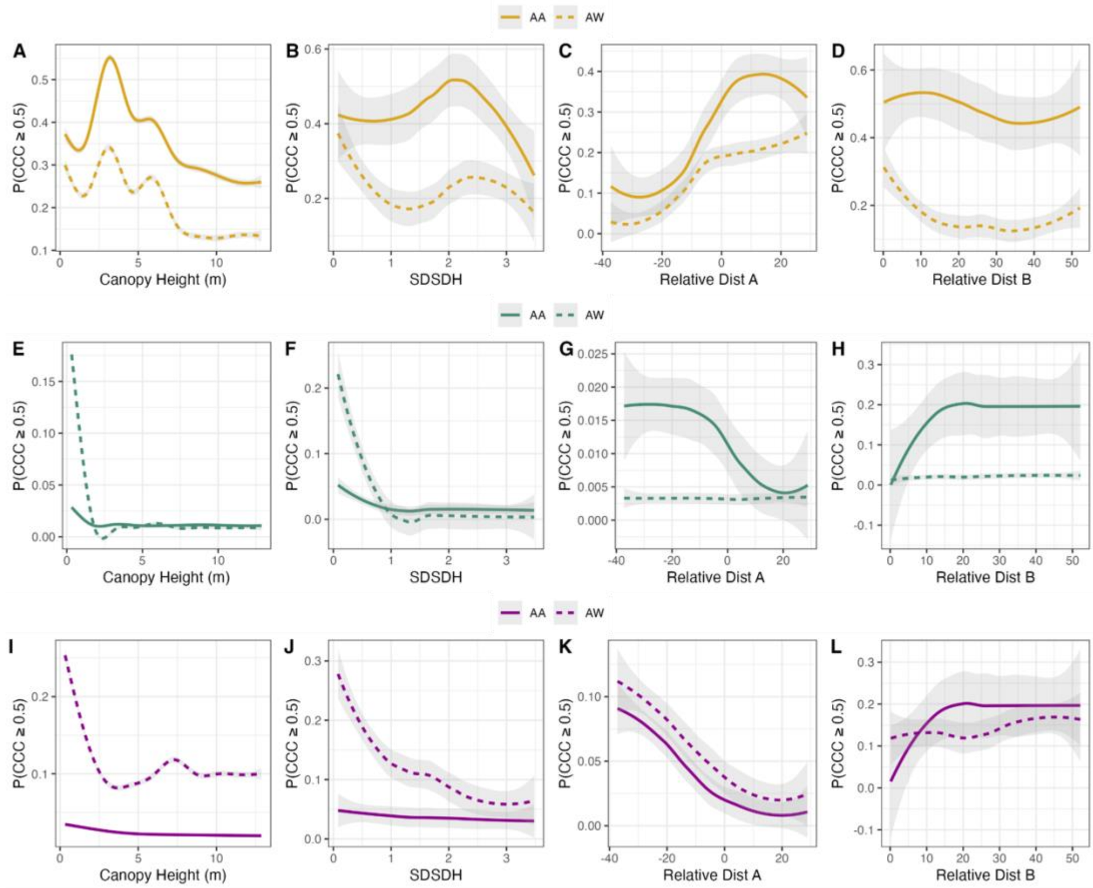


Figure 5: Sensitivity analysis for the model estimating the probability of a sampling height pair having a concordance correlation coefficient (CCC ≥ 0.5) for the MBR (A-D), the AE (E-H), and WP (I-L) by canopy level (AA, AW) for the mean canopy height (m), the standard deviation of canopy height (SDSDH), and the distance between the measurement height and the canopy at the upper (Relative Dist A) and lower (Relative Dist B) measurement levels influenced the reliability of sampling height pairs, respectively.

Sensor placement (z) relative to the canopy (h_c) had important effects on the concordance of GFs for CO_2 (Figure 6). The upper sensor required clearance above the canopy (~5 m), and wider sensor separation amplified the gradient signal (r = 0.34 for MBR). Canopy structure also had important implications for concordance. Short, open canopies had higher concordance, and CCC correlated negatively with canopy height, VAI, FHD, VCI, and top rugosity, and positively with gap fraction (DGF, GFP). Tall, dense, vertically complex canopies (large rugosity, high FHD) pushed profiles into the roughness sublayer, resulting in lower CCC. Of the three approaches explored, the MBR was the most sensitive to favorable conditions, increasing when sensor placement and canopy structure were ideal (mean CCC = 0.39 in AA pairs, peaking near 0.96 at UNDE and short-grass sites like KONZ/STER). The AE and WP plateau at much lower levels (mean ≈ 0.10 – 0.12) across canopies, suggesting that their flux estimates were limited by the underlying eddy-diffusivity parameterization rather than by sensor placement. While patterns for H_2O were similar, cross-canopy fluxes for H_2O showed stronger alignment with EC (Figure A11).

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

Open Access Earth System Science Data Discussions

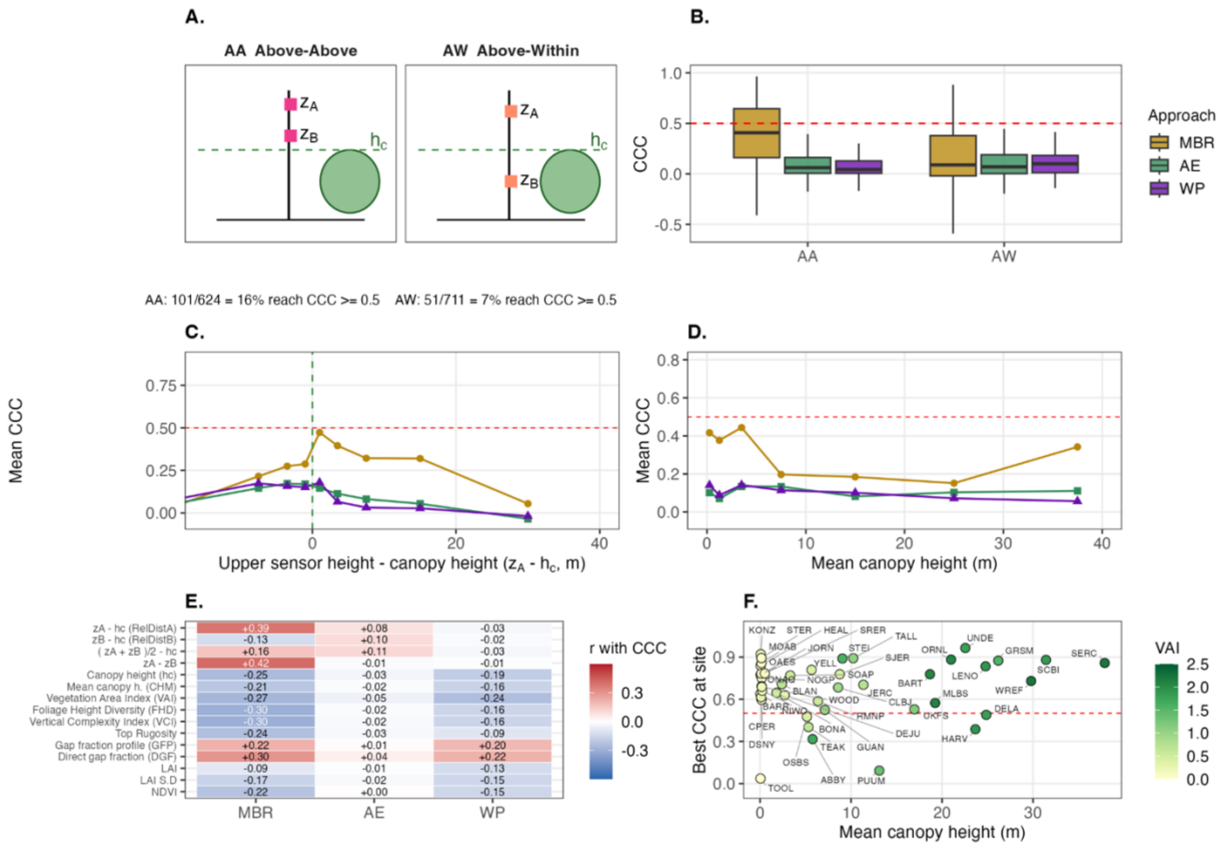


Figure 6: Sensor placement (z_A) relative to the mean canopy height (h_c; A) and its influence on the concordance (CCC) at NEON sites (B) for CO_2. The relative distance ($z_A - h_c$) versus mean CCC by approach (C) and the canopy height (h_c) versus mean CCC by approach (D). The correlation between canopy geometry and CCC (E), and the highest CCC at each site for CO_2 relative to hc, and the vegetation area index (VAI; F) indicate the importance of canopy structure. Red dashed lines mark a CCC ≥ 0.5, and green dashed lines mark the mean canopy height.

3.3 Gradient Flux Ensemble

Temporal overlap in GFs from different sampling height pairs and approaches was common across sites and ranged from 0% to 48% for sampling height pairs with a CCC ≥ 0.5 (Figure A12). The highest overlap was between the AE and WP, whereas data from all three approaches overlapped for fewer than 5% of available fluxes on average (Figure A12). There were small differences in ensemble method performance (Table 4). Therefore, we used the mean ensemble in all analyses. The concordance between ensembled GFs and EC was just as strong as the mean CCC across sites (Figure 7), suggesting that the GF ensemble did not reduce concordance with EC.

Table 4: Comparison of ensemble methods for GF with a CCC ≥ 0.5. Methods included Bayesian fusion, inverse RMSE-weighted, approach-weighted, mean, median, and weighted median uncertainty.

Ensemble Method	Mean R^2	S.D. R^2	Mean CCC	S.D. CCC	Mean RMSE	S.D. RMSE

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

Bayesian fusion	0.501	0.12	0.66	0.09	3.44	3.2
Inverse RMSE weighted	0.501	0.119	0.66	0.09	3.46	3.2
Approach weighted	0.499	0.118	0.66	0.08	3.47	3.2
Mean	0.499	0.117	0.66	0.08	3.49	3.2
Median	0.496	0.116	0.66	0.08	3.49	3.2
Median weighted	0.49	0.12	0.65	0.09	3.47	3.2

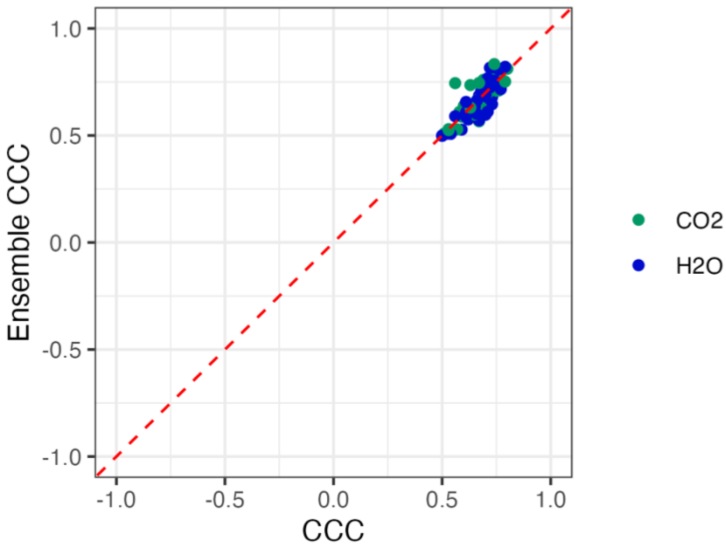


Figure 7: A comparison of the concordance between the mean CCC for reliable sampling height pairs at NEON tower sites and the ensemble GF CCC.

3.4 Ecological Assessment of Gradient Fluxes

Site-level comparisons between the ensemble GF and EC fluxes showed moderate agreement, with substantial variation among sites and somewhat stronger consistency in temporal patterns than in distributional spread (Figures A13-A16). In the one-to-one comparisons, the site-level R^2 averaged 0.52 for CO_2 and 0.50 for H_2O, indicating that ensemble GF fluxes generally captured a meaningful proportion of site-level variability in EC fluxes (Figures A15-A16). Distributional comparisons showed that GF fluxes were typically more variable than EC fluxes within sites, particularly for CO_2. The standard deviation of GF exceeded that of EC at 35 of 38 CO_2 sites and 31 of 44 H_2O sites, with median GF: EC standard deviation ratios of 1.51 and 1.24, respectively. Mean differences between methods were generally modest but were larger for CO_2 than for H_2O, suggesting that the ensemble approach reproduced central tendency more closely for H_2O than for CO_2. Together, these results indicate that the ensemble GF estimates reproduced flux dynamics across sites, but often with a wider flux distribution than EC, especially for CO_2.

Seasonal diel curves had strong correlations ($R^2 \approx 0.58$ - 0.99) with EC curves for both CO_2 (n = 31) and H_2O (n = 43; Figure 8). Most towers produced valid CO_2 diels for at least one season, and most sites (64% for CO_2 and 89% for H_2O) had data for the spring and summer curves. Although the GF tended to undersample mid-day CO_2 measurements (Figure A5B) and early-

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

morning H_2O measurements (Figure A5C), there were no consistent patterns across ecosystem types in over- or under-sampling of total daily diel curves. Seasonal daily flux rates differed by a mean absolute value of ~ 2.1 g CO_2 m^{-2} day^{-1} and were highly variable (2.9 S.D.) across sites and seasons. The H_2O datasets were larger, and correlations between GF and EC were stronger than those for CO_2. Seasonal daily flux rates differed by a mean absolute value of ~ 202 g H_2O m^{-2} day^{-1} (181 S.D.) across sites and seasons. Although differences in total daily fluxes across seasons yielded similar implications for the magnitude and sink/source strength of ecosystem fluxes, biases across ecosystem types ranged from 30 to 50% for CO_2 and 1 to 40% for H_2O. Correlations between GF and EC seasonal diel curves indicate that forested ecosystems were more likely to have lower R^2, and biases were larger for CO_2 than they were for H_2O. Altogether, ensemble GFs provided ecosystem assessments that aligned with the EC method, and performance limitations were influenced by the availability of sufficient data to capture diel patterns across seasons. Site-based diel curves are shown in Supplement A (Figures A17 and A18), and comparisons of GF and EC-based diels are shown in Figure A19 and Table A2.

Figure 8: The correlation strength (R^2) and slope (y) for total daily ensemble GF versus EC (A) based on diel curves for CO_2 (g CO_2 m^{-2} day^{-1}; B) and H_2O (g H_2O m^{-2} day^{-1}; C) across seasons for ecosystem types. The dashed line in panel A is a one-to-one line. Panels B and C show diel curves for a subset of sites: the arid grassland in the Jornada Basin (JORN), the tallgrass prairie in Konza Prairie (KONZ), the tropical evergreen forest at Guanica Forest (GUAN), and the temperate mixed forest at Harvard Forest (HARV). The solid lines in panels B and C are GF diels, and the dashed lines are EC diel curves. Panels include only sites with sufficient data to fit diel curves.

4 Discussion

Benchmark analyses across a continental network of 47 flux towers show that the GF method can recover ecosystem-scale CO_2 and H_2O exchange rates. Alignment between the GF and EC methods was strongest in structurally simple ecosystems and declined in complex canopies. Although temporal sampling was limited, an ensemble approach improved temporal coverage by combining all reliable estimates across approaches and sampling-height pairs. These findings suggest that the GF method can be used as a strategic tool to complement EC in multi-gas observing networks. Rather than asking whether GF methods can universally replace EC, our results identify where and when they provide operationally useful information at the ecosystem scale. The broader significance of this result extends beyond CO_2 and H_2O. Researchers increasingly seek to quantify gases for which fast-response EC instrumentation remains expensive, technically demanding, or unavailable. In such cases, profile-based approaches that leverage existing tower infrastructure offer a practical pathway to expand atmospheric monitoring capacity.

4.1 How well does the GF method perform across different ecosystems?

Across terrestrial ecosystems, the GF method produced reliable estimates of ecosystem fluxes (Meredith et al., 2014; Mölder et al., 1999; Raupach and Legg, 1984; Zhao et al., 2019). Ecosystems with short canopies often showed stronger agreement

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

between GF and EC compared to those with tall canopies. Lower performance in tall ecosystems was likely influenced by the proximity of tall canopies to the roughness sublayer, where the validity of similarity theory is in doubt (Simpson et al., 1997). With tall canopies, the distance between measurement pairs can also become limited by the tower height and gradient strength (Mölder et al., 1999; Simpson et al., 1998). Flux-gradient theories have been found to overestimate scalar fluxes within the roughness sublayer (ground to 2 or 3 times the canopy height) because turbulent structure is strongly influenced by the canopy (Baldocchi and Meyers, 1991; Högström et al., 1989; Kenion et al., 2024; Raupach and Thom, 1981; Simpson et al., 1998), especially for tall vegetation (Garratt, 1978). That said, the application of the GF method requires appropriate measurement systems and can benefit from site-specific universal functions to improve GF performance within the roughness sublayer (Mölder et al., 1999), especially above forests, at just 1.2-1.4 times the canopy height (Meredith et al., 2014; Simpson et al., 1998).

4.2 Which sampling height pairs provide the most reliable GFs?

The concordance of reliable sampling height pairs varied across GF approaches, suggesting that optimal measurement heights differ by approach and by canopy structure. For short canopies ($h_c \lesssim 5$ m), all GF approaches had stronger concordance when the lower sensor was at the canopy top (±2 m), and the upper sensor was 5–10 m higher. Concordance peaked at 0.74 (CO_2 MBR) and 0.81 (H_2O MBR). For tall canopies ($h_c \gtrsim 15$ m), the gas source distribution matters. For CO_2, a z_A near or just below the canopy top (−5 to +5 m) and a z_B at about 10–20 m above produced the highest concordance for the MBR (CCC ≈ 0.58). For H_2O, the MBR showed higher concordance when both measurements were above-canopy, whereas AE and WP performed better when z_A was within the canopy. The best measurement pairs had a z_A at $h_c - 15$ to $- 20$ m (trunk space) and z_B at $h_c +$ 10–15 m. The AW configuration boosted the AE and WP from ~0.18 to ~0.64 in tall forests, allowing the gradient to straddle the transpiration source rather than sampling well-mixed air AA. Differences between gases resulted from the source distribution. With a single dominant source (transpiring leaves at the canopy crown), the steepest gradient for H_2O was across the canopy. With competing sources/sinks (uptake at leaves and respiration in the soil), CO_2 showed the strongest gradient in the constant-flux layer. The MBR benefited from above-canopy clearance for both gases.

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

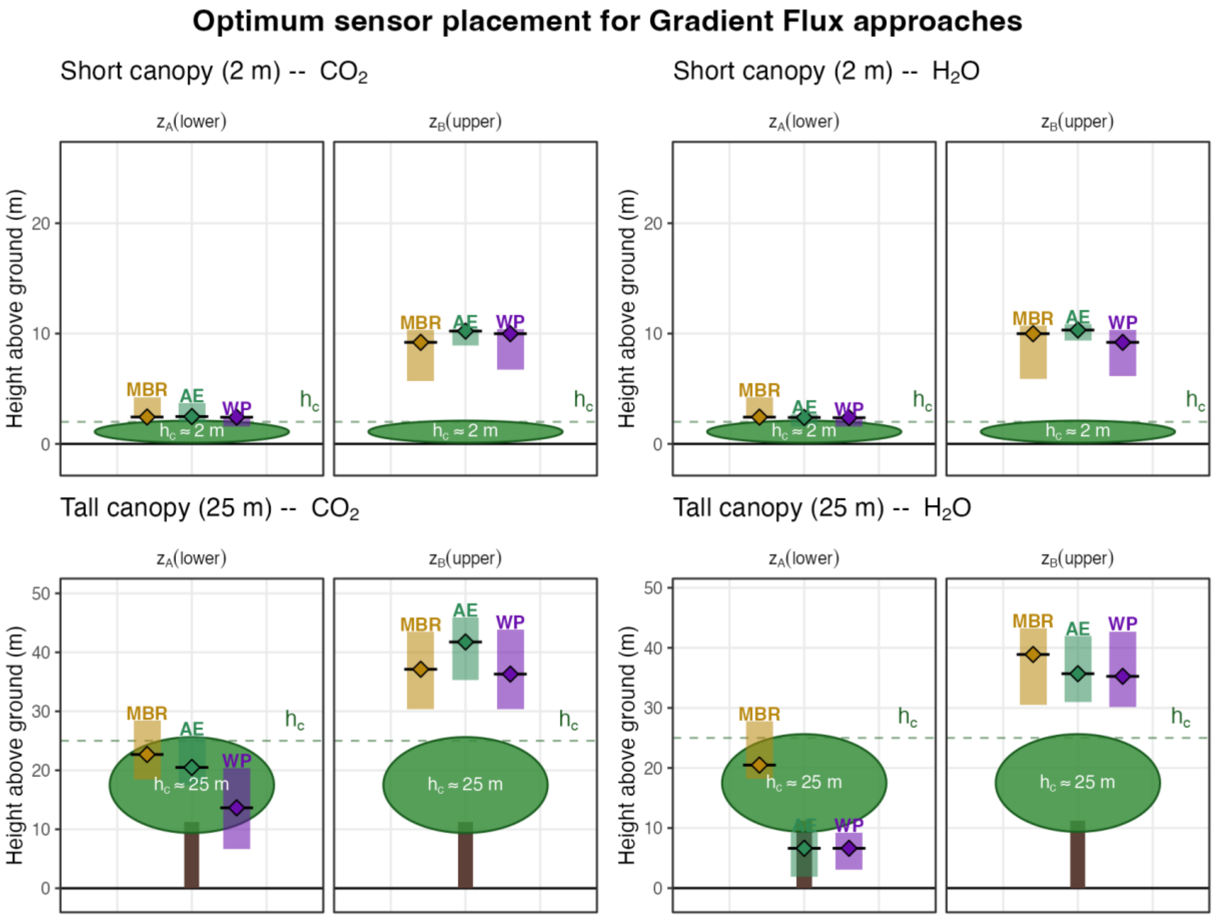


Figure 9: The optimum sensor placement relative to the canopy for the gradient flux approaches (MBR, AE, WP). The colored band is the interquartile range (IQR) for the lower (z_A) and upper (z_B) measurement heights across the top-quartile CCC pairs, with a diamond at the median. The green dashed line is the mean canopy height (hc).

4.3 Does the GF method provide a reliable ecological assessment?

While the GF method yielded reliable ecosystem fluxes across different ecosystems and seasons, significant data loss due to filtering and the extraction of reliable sampling height pairs for these approaches can limit its utility. However, sparse observations contained an important and previously underused feature: complementarity. Periods rejected by one approach were frequently retained by another. By combining reliable estimates across approaches and configurations, the ensemble framework converted fragmented half-hourly observations into coherent diel and seasonal records. Ensemble GF fluxes preserved concordance with EC while markedly improving temporal representativeness. This is the first study to show that ensemble GFs could be used to understand patterns in ecosystem fluxes across seasons. Strong alignment between the EC and ensemble GF diel curves suggests that this approach can be used to understand patterns in productivity, respiration, and transpiration. While the GF method is unlikely to provide the same temporal resolution of ecosystem dynamics as the EC

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

Open Access Earth System Science Data Discussions

method, the GF approach can yield sufficient data to understand diel patterns and the temperature- and moisture-sensitivity of ecosystem fluxes. This result has broader implications for environmental observing systems. Many ecological datasets are incomplete, irregular, and method-dependent. Ensemble frameworks can exploit partial redundancy across measurement approaches to recover useful signals that no individual method can provide alone. In the present study, ensemble GF estimates reproduced diel patterns of CO_2 and H_2O across diverse ecosystems, demonstrating that high-frequency completeness is not always necessary to recover ecologically meaningful dynamics.

4.4 Limitations of the GF method

The GFs included in this study had the well-documented limitations of the GF method itself (Maier and Schack-Kirchner, 2014): the approach was biased toward periods and ecosystems with large concentration gradients and performed poorly when true fluxes were near zero (Laubach et al., 2016; Wu et al., 2015; Zhao et al., 2019). Since no independent method for identifying near-zero fluxes was applied here, GF estimates in this dataset may be biased toward periods with detectable, non-zero fluxes. Even with these limitations, the GF method remained beneficial when instrumentation was limited, footprints were small (Zhao et al., 2019), and there was interest in decomposing ecosystem fluxes by source contribution (i.e., soil and canopy fluxes).

A related limitation of this dataset is its fixed sampling protocol. When measurements between two heights alternate quickly, the gradient is calculated within similar timeframes, reducing low-frequency noise and errors due to low-frequency drifts in the analyzer (Wagner-Riddle et al., 2005). While switching times for the GF method can be as fast as 4 seconds (Wagner-Riddle et al., 1996), 10 seconds (Edwards et al., 2003), or 30 seconds with lower-power pumps (Brown et al., 2024) shorter switching times are associated with declines in sensor precision, leading to a trade-off between the long-time needed to achieve the desired precision and the short time needed to prevent major temporal changes in the concentration of the trace gas. As implemented here, the 6-minute continuous protocol likely introduced more low-frequency noise and drift-related error into these GF estimates than a faster-switching design would have, particularly at the upper canopy heights where the GF method is most often applied.

Poor agreement between EC and GF at different sampling heights was occasionally due to opposite signs between GF and EC, a pattern we call "counter-gradient.” Counter-gradient fluxes occur when the transport of a quantity is in the opposite direction to its mean gradient, contradicting the standard gradient-diffusion theory. Counter-gradient fluxes are primarily important within the canopy. Across NEON sites, counter-gradients were more frequent at taller-canopy sites and influenced which sampling height pairs were reliable. When counter-gradient fluxes were observed within the surface layer, this was a likely indication of an issue with the measurement approach rather than a physical phenomenon itself. For example, when switching flow lines, using buffer volumes can improve temporal representativeness, whereas insufficient buffering can lead to such artifacts. Long delays between sampling levels also increase sensitivity to non-stationarity that can lead to counter-gradients. While filtering out counter-gradients led to significant data loss, the remaining data still allow canopy measurements to produce

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

reliable ecosystem GFs, as found in other studies (Meredith et al., 2014; Wu et al., 2015). In addition to counter-gradient flows, gradient strength was an important factor in our ability to use the GF method. Gradient strength, governed by the vertical spacing between sampling heights, was the other main constraint on which height pairs produced reliable fluxes in this dataset (Cellier and Brunet, 1992; Meredith et al., 2014). Balancing the distance between pairs in the presence of canopy interference is a challenge in ecosystems with tall canopies (Cellier and Brunet, 1992; Meredith et al., 2014).

5 Conclusions

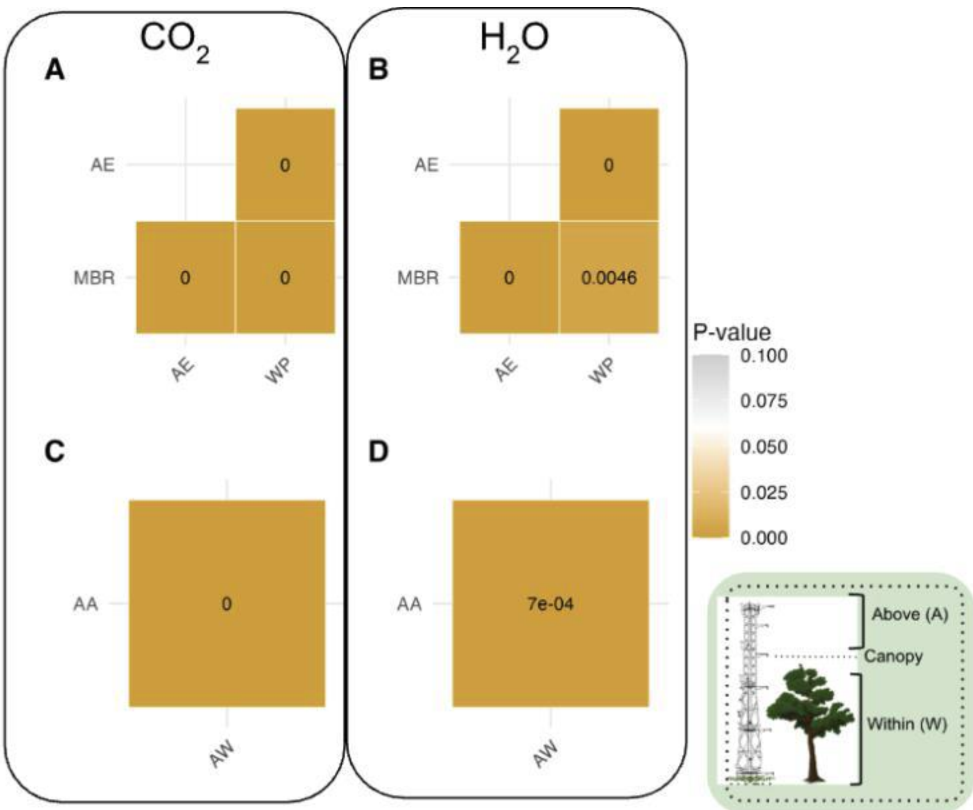
Flux science is entering a multi-gas era in which carbon, nitrogen, water, and atmospheric chemistry must be observed simultaneously. No single method will satisfy all observational needs. Benchmark analyses show that strategically deployed GF approaches can complement EC networks by expanding the number of measurable gases, sites, and environmental conditions that existing tower infrastructure captures. With the GF method, it is valuable to have a system that can leverage multiple GF approaches. The NEON tower network design enabled us to use independent flux measurements to determine when and where we can apply the GF method to measure ecosystem fluxes. Benchmark analyses suggest several design principles for integrating the GF method into existing networks. First, prioritize short or moderately tall canopies where above-canopy gradients are measurable. Second, where forests are targeted, maximize the height of upper inlets and evaluate cross-canopy configurations rather than assuming a single optimal pair. Third, deploy multiple methods simultaneously where feasible to enable ensemble recovery. In this sense, the future of GF methods is likely to be network-informed. Future work should focus on canopy corrections for tall/complex forests, approaches to determine near-zero fluxes, decomposing ecosystem fluxes by source contribution, and developing standardized data products. Beyond gradient-flux evaluation, the benchmark dataset can support (1) machine-learning flux emulators, (2) testing canopy transport parameterizations, (3) sensor placement optimization, (4) uncertainty benchmarking for emerging trace-gas methods, and (5) educational training datasets for micrometeorology.

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

Appendix A. Supplementary Tables and Figures

600 **Table A1. Evaluation of operational concordance thresholds for defining reliable sampling-height pairs (RSHP). We compare the percentage of sampling height pairs retained, the mean sign agreement, the mean flux bias, the absolute flux bias, the correlation, and the root-mean-square error (RMSE) between GFs and EC.**

Gas	Threshold	Sampling height pairs retained (%)	Sign agreement	Flux bias	Absolute flux bias	Correlation	RMSE
CO2	CCC >= 0.3	69.2	0.837	0.074	3.996	0.770	0.292
CO2	CCC >= 0.5	33.8	0.858	-0.146	3.046	0.825	0.249
CO2	CCC >= 0.7	12.7	0.876	-0.203	2.569	0.891	0.194
H2O	CCC >= 0.3	81.9	0.854	-0.007	0.858	0.831	0.182
H2O	CCC >= 0.5	57.3	0.860	-0.046	0.670	0.873	0.150
H2O	CCC >= 0.7	23.8	0.883	-0.039	0.513	0.933	0.106



605 **Figure A1. Pairwise Wilcox test for differences in CCC by GF approach (MBR, AE, and WP) for CO_2 (A) and H_2O (B), and by canopy level pairs (AA and AW) for CO_2 (C) and H_2O (D).**

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

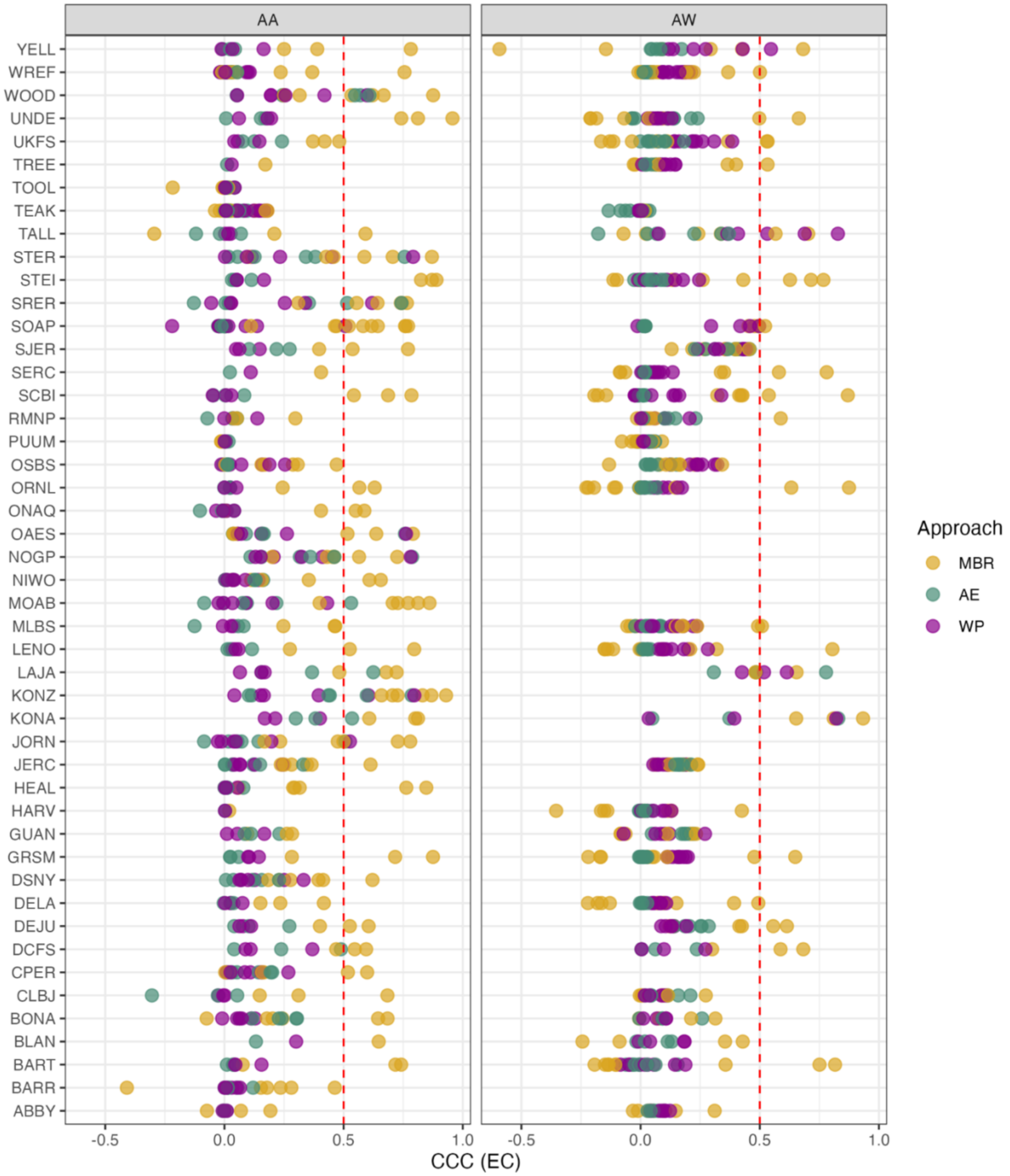
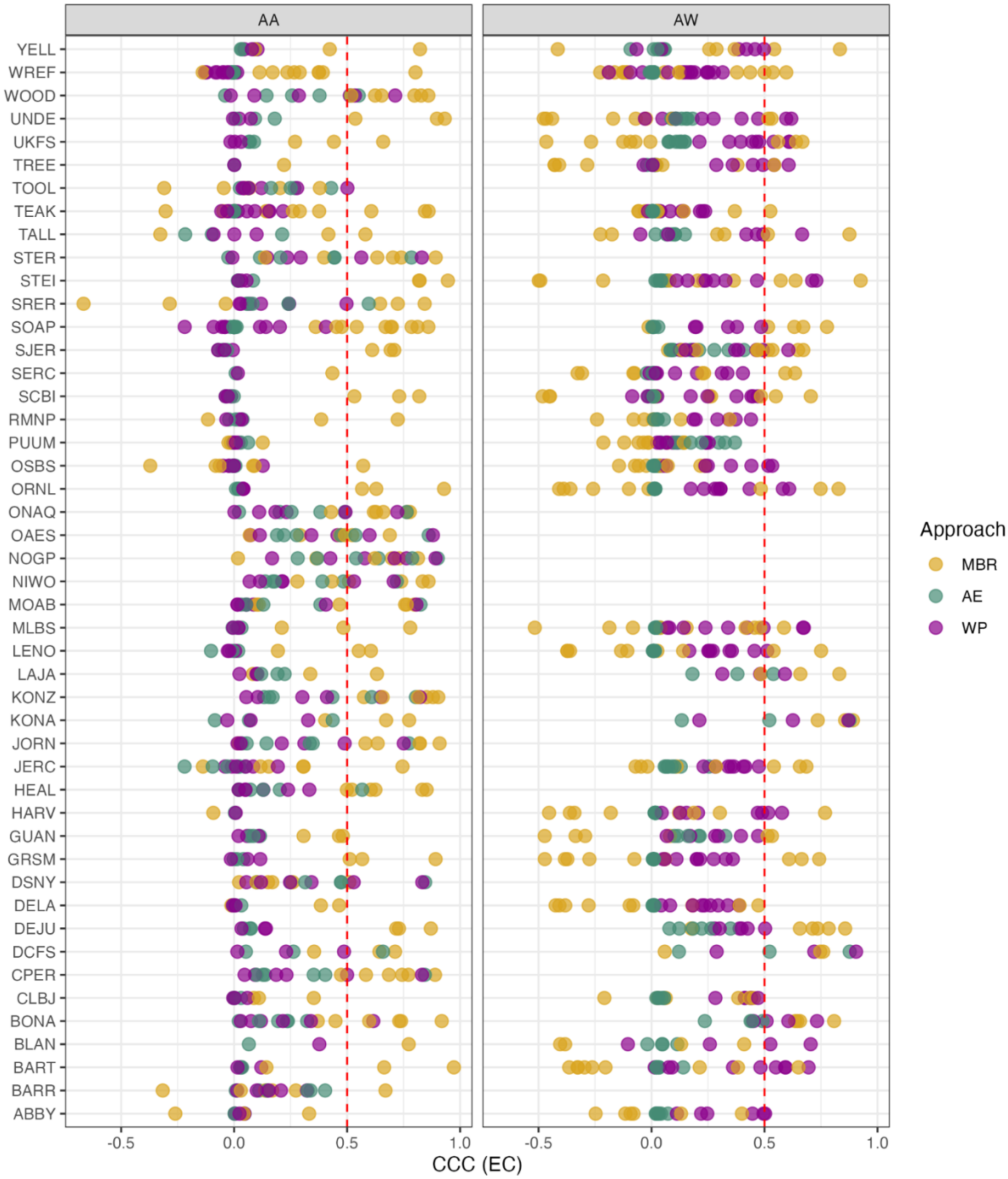


Figure A2. The CCC values for the GF approach (MBR, AE, and WP) and canopy level (AA and AW) for CO_2 across NEON sites. The vertical dashed lines show a threshold for CCC ≥ 0.5.

Open Access Earth System Science Data Discussions



610 **Figure A3. The CCC values for the GF approach (MBR, AE, and WP) and canopy level (AA and AW) for H_2O across NEON sites. The vertical dashed lines show a threshold for CCC ≥ 0.5.**

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

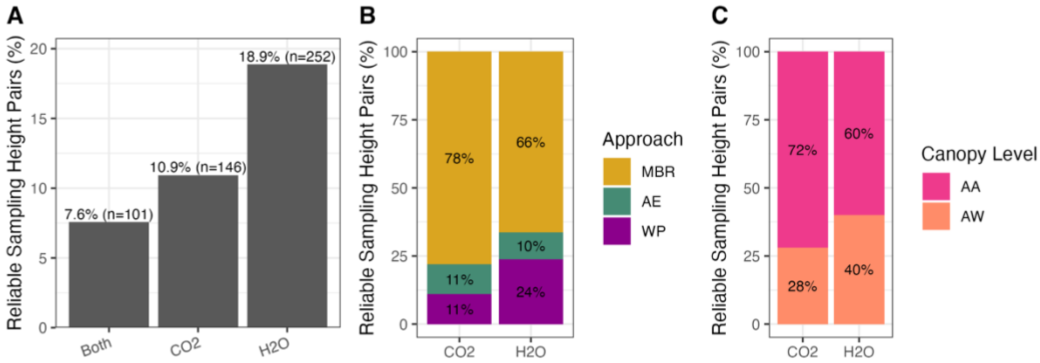


Figure A4. The percentage of reliable sampling height pairs (CCC > 0.5) for CO_2 and H_2O (A), by approach (B), and canopy level (C).

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

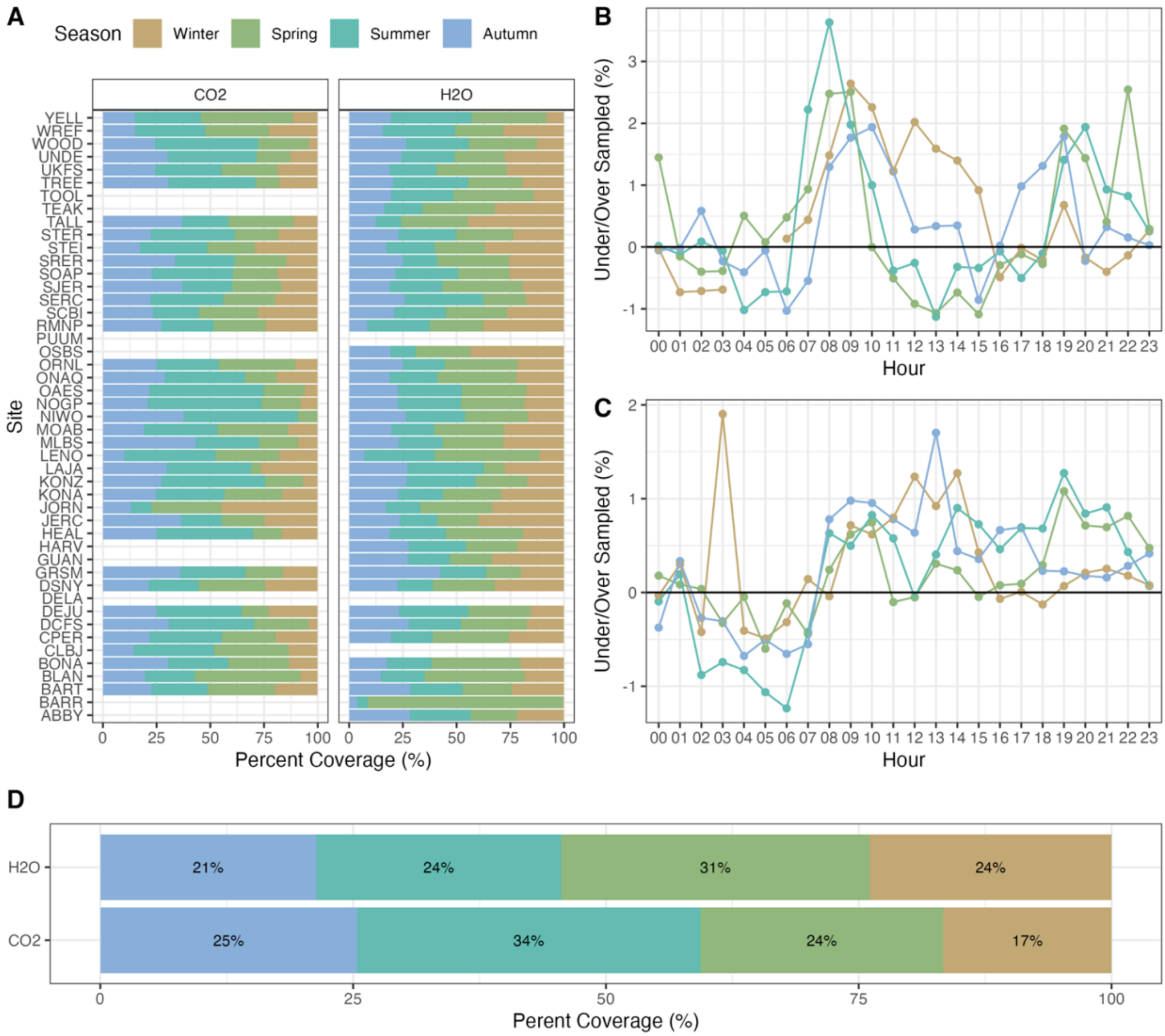


Figure A5. The percent coverage of gradient flux (GF) data from reliable sampling height pairs (CCC > 0.5) across sites by season (A) and hour for CO_2 (B) and H_2O (C). The summary across all sites (D) indicates that the data is skewed towards the summer for CO_2 and spring for H_2O.

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

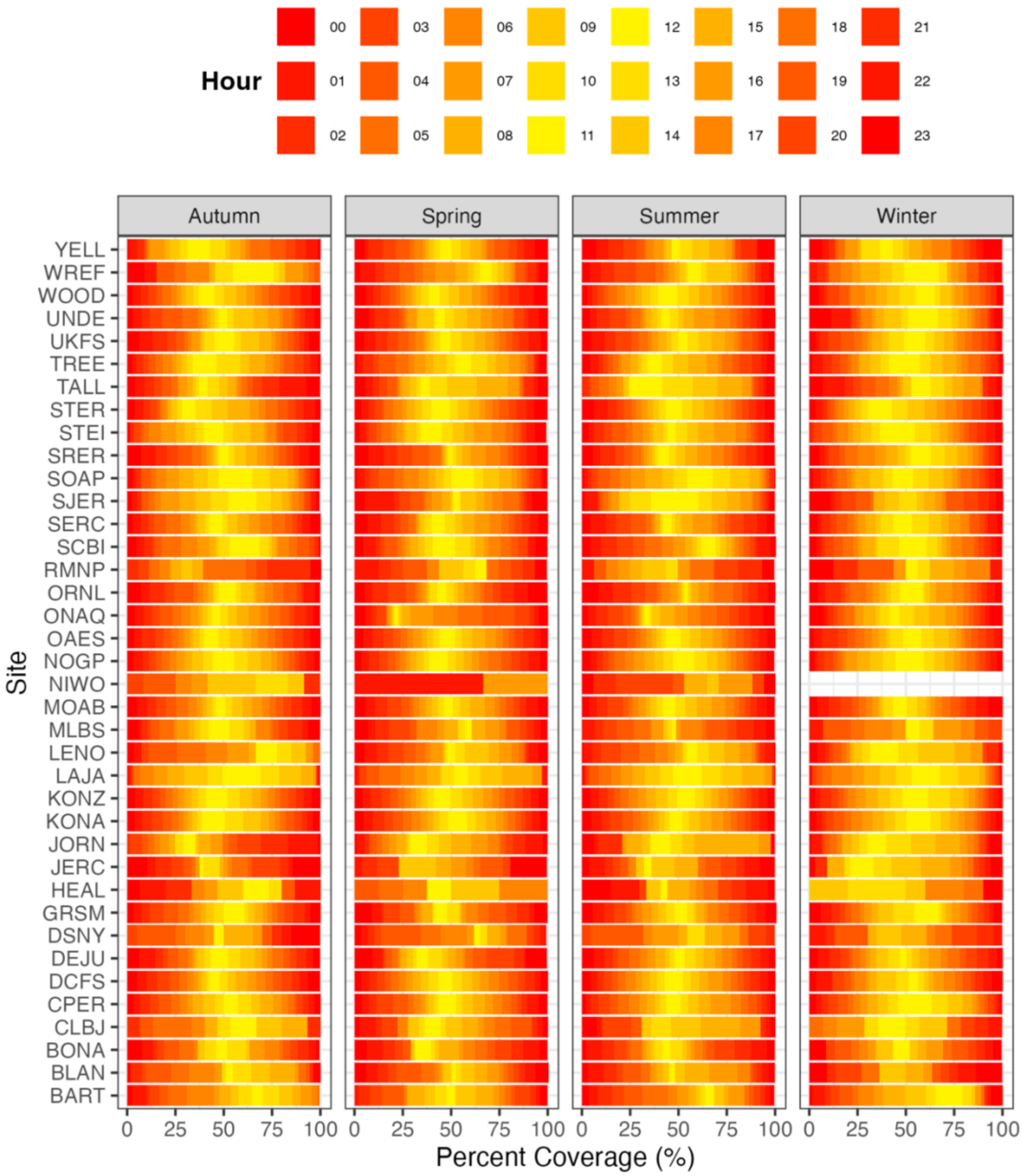


Figure A6. The percent coverage of gradient flux CO_2 data by hour for each site with a CCC $\geq$ 0.5.

625

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

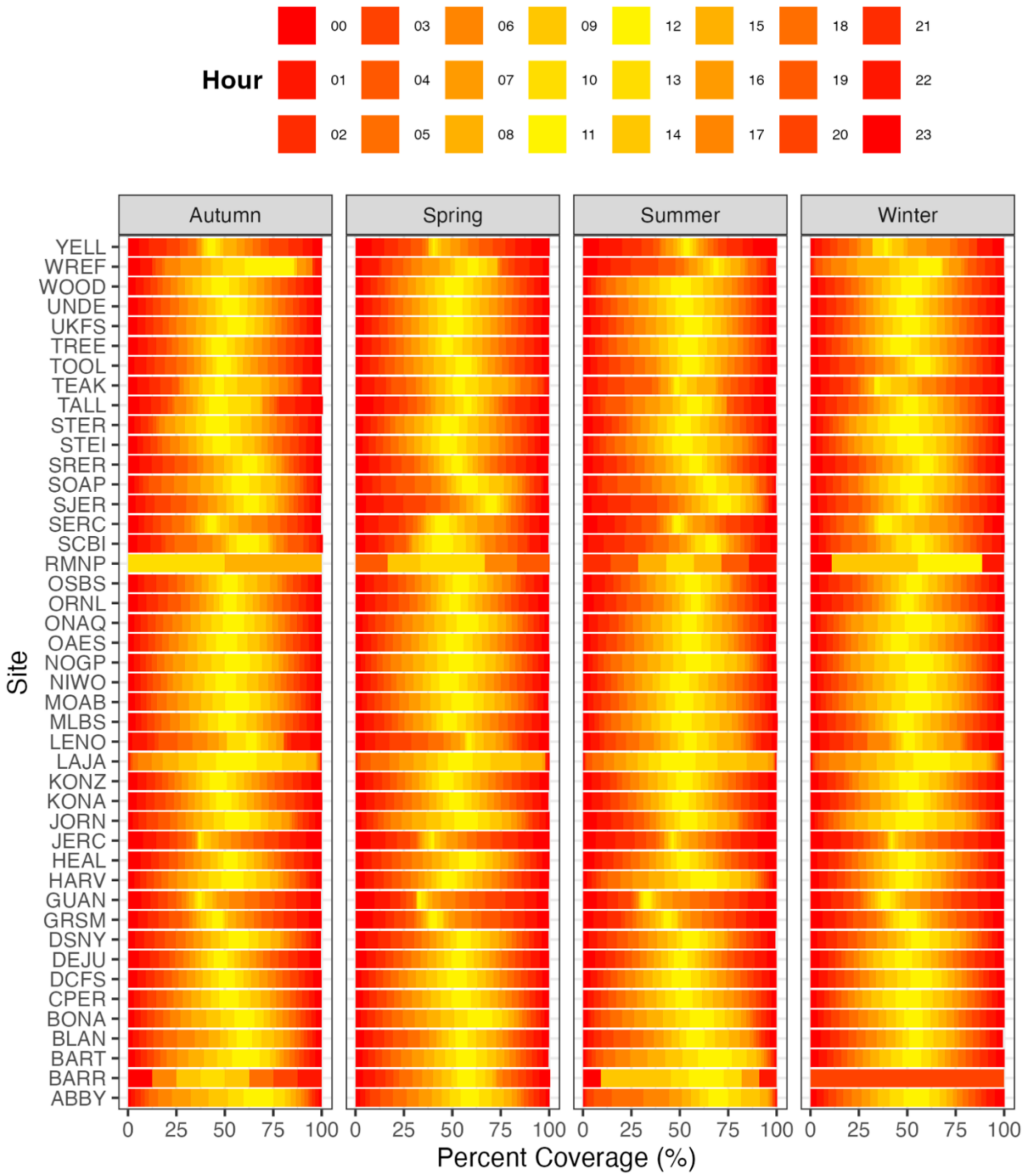
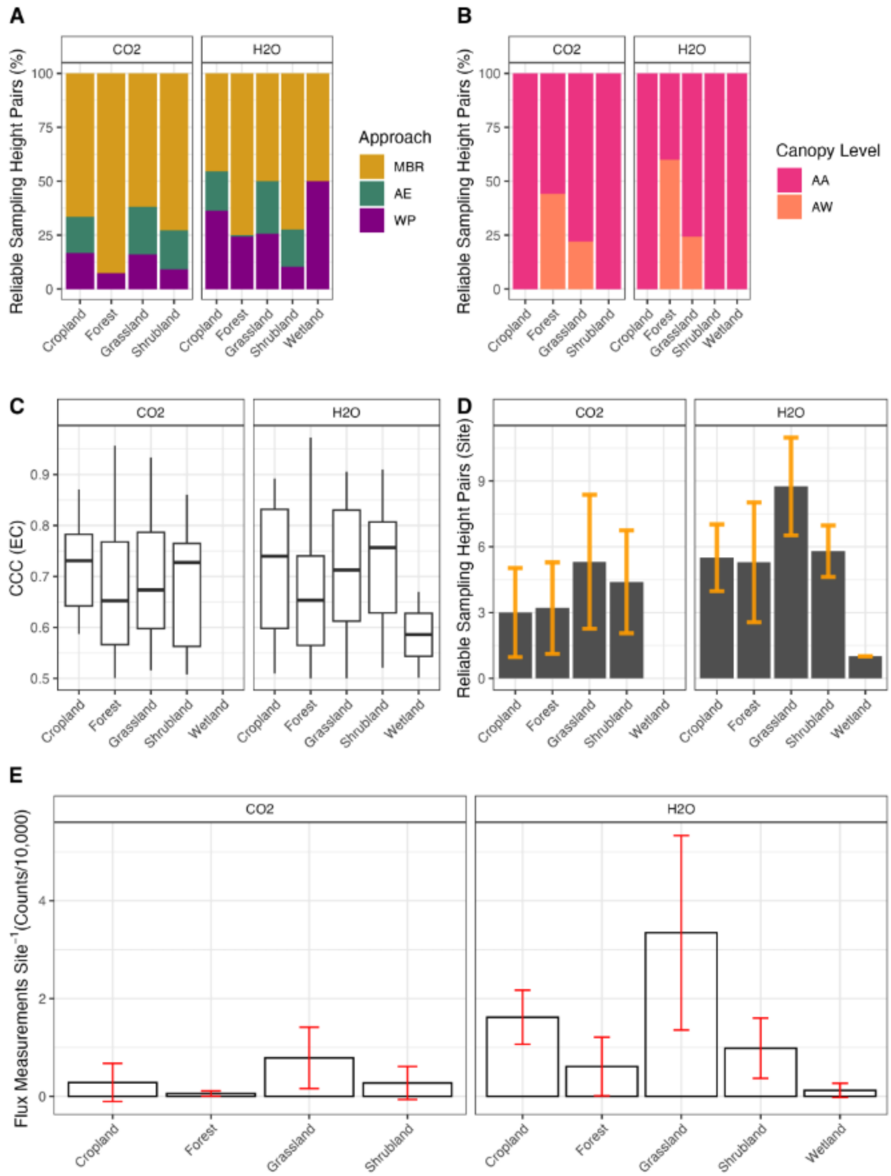


Figure A7. The percent coverage of gradient flux H_2O data by hour for each site with a CCC ≥ 0.5.

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

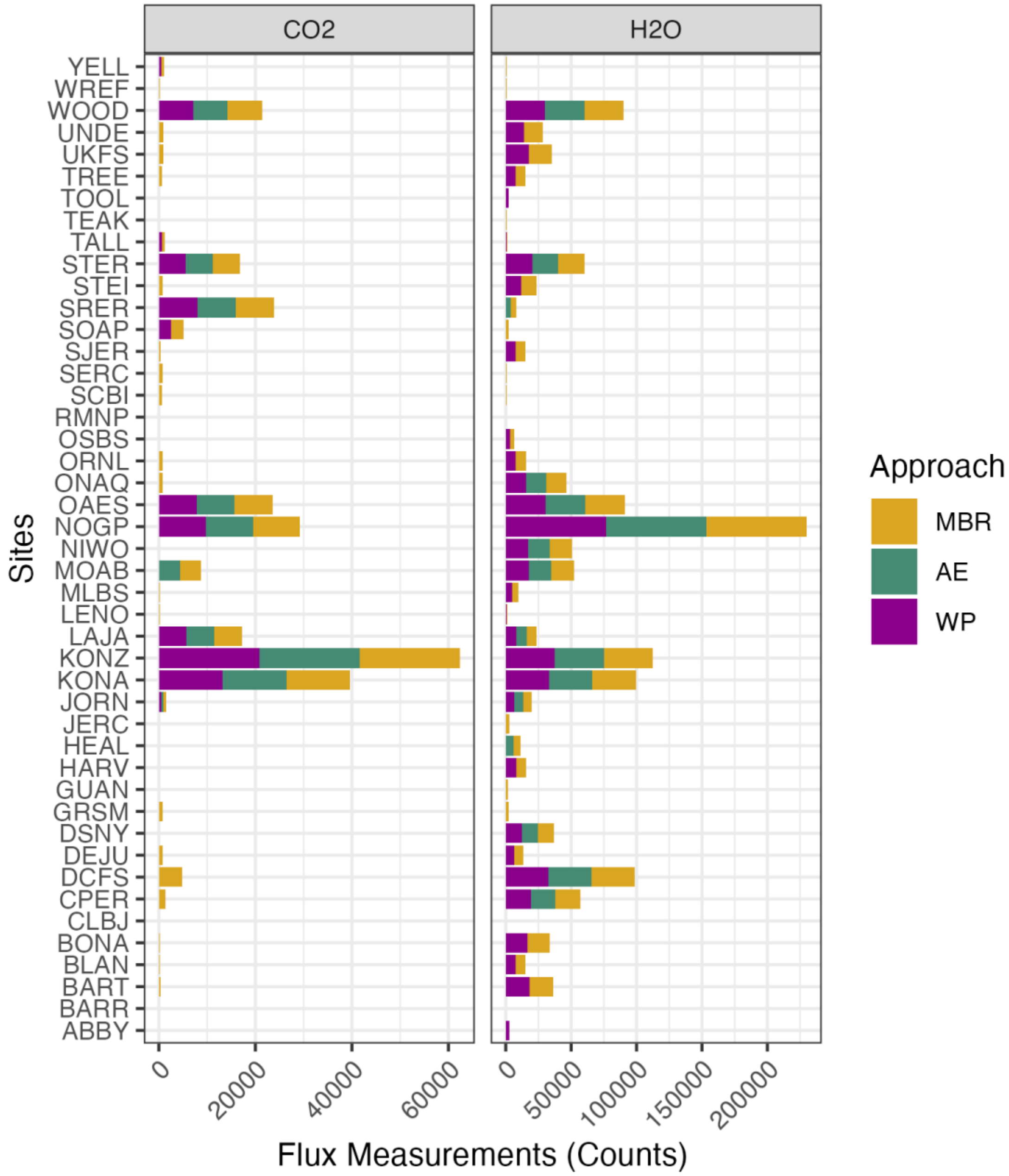
Open Access Earth System Science Data Discussions



630 **Figure A8. The percentage of reliable sampling height pairs (CCC > 0.5) for CO_2 and H_2O by approach (MBR, AE, and WP; A), canopy level (AA and AW; B), and ecosystem type. The concordance (CCC) varied across ecosystems (C), resulting in variation in the average number of reliable sampling height pairs (D) and the average size of the GF dataset at tower sites (E).**

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

Open Access Earth System Science Data Discussions



635 **Figure A9. The number of flux measurements at each site by the approach (MBR, AE, WP) with a CCC ≥ 0.5.**

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

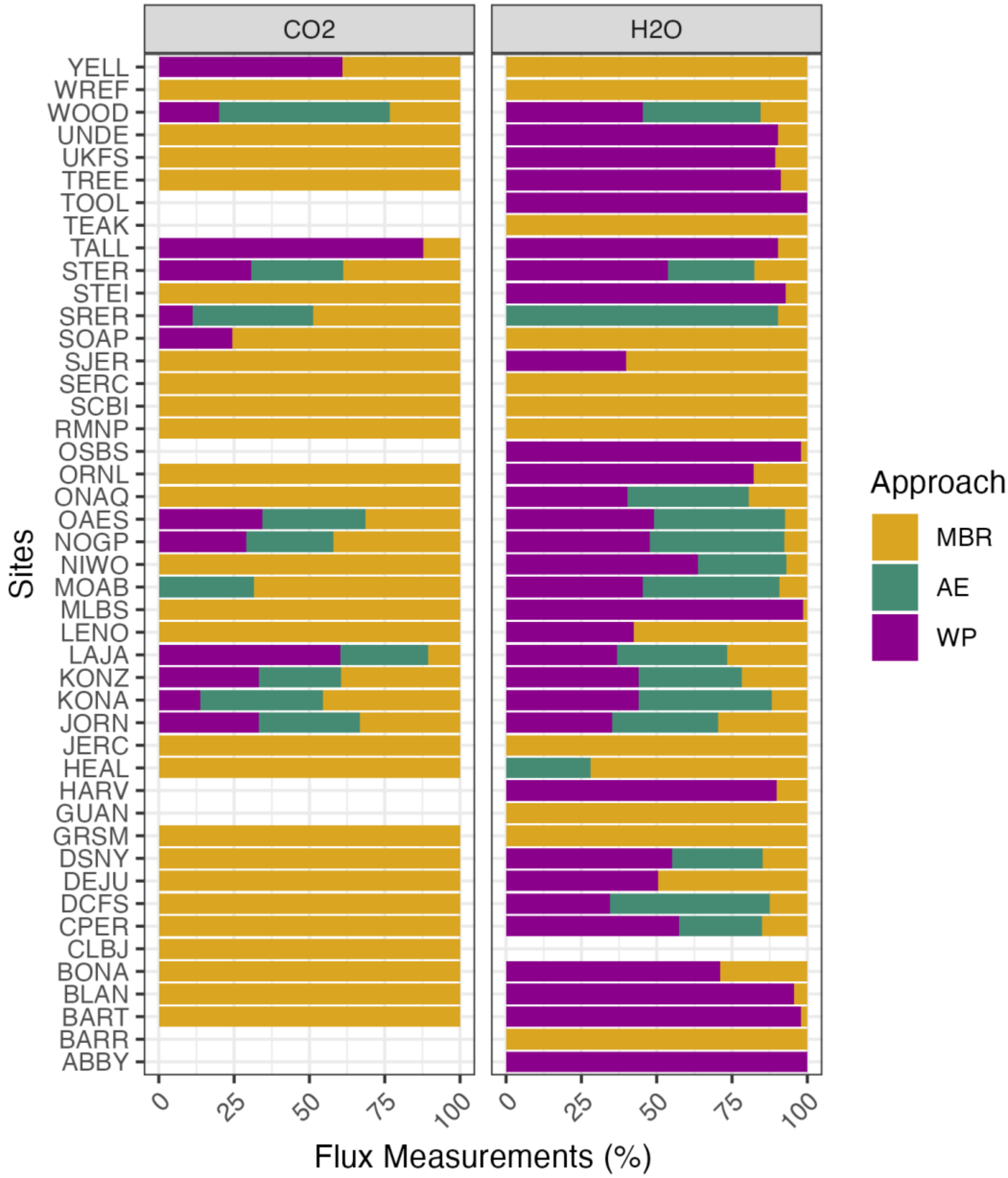


Figure A10. The percent of flux measurements at each site by the approach (MBR, AE, WP) with a CCC ≥ 0.5.

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

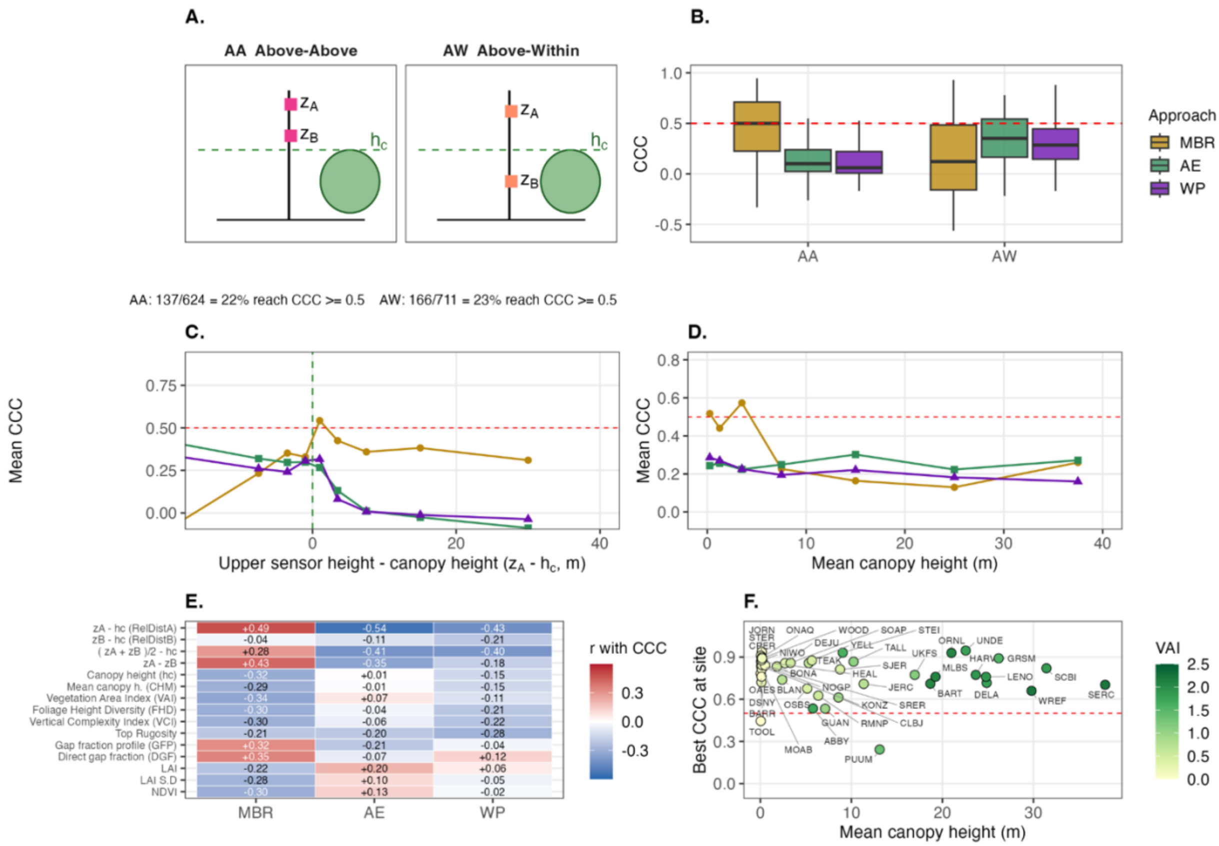


Figure A11. Sensor placement (z_A) relative to the mean canopy height (h_c; A) and its influence on the concordance (CCC) at NEON sites (B) for H_2O. The relative distance ($z_A - h_c$) versus mean CCC by approach (C) and the canopy height (h_c) versus mean CCC by approach (D). The correlation between canopy geometry and CCC (E), and the highest CCC at each site for H_2O relative to h_c, and the vegetation area index (VAI; F) indicate the importance of canopy structure. Red dashed lines mark a CCC ≥ 0.5, and green dashed lines mark the h_c.

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

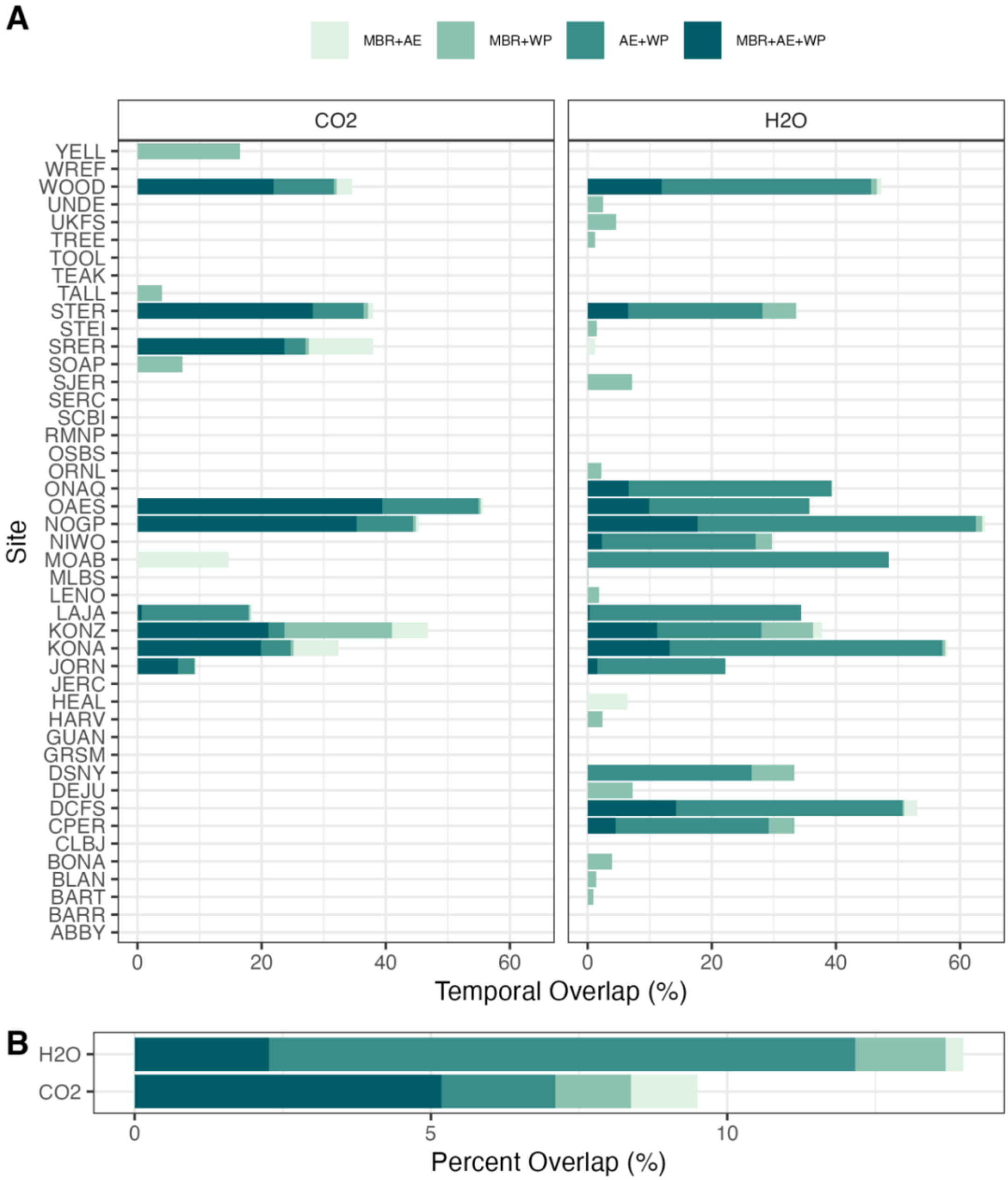


Figure A12. Temporal overlap in GF with a CCC ≥ 0.5 across sites and approaches (A) indicates that different GF approaches are essential for temporal coverage at many sites. Average temporal coverage across sites is higher for H_2O than for CO_2 (B).

650

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

Open Access Earth System Science Data Discussions

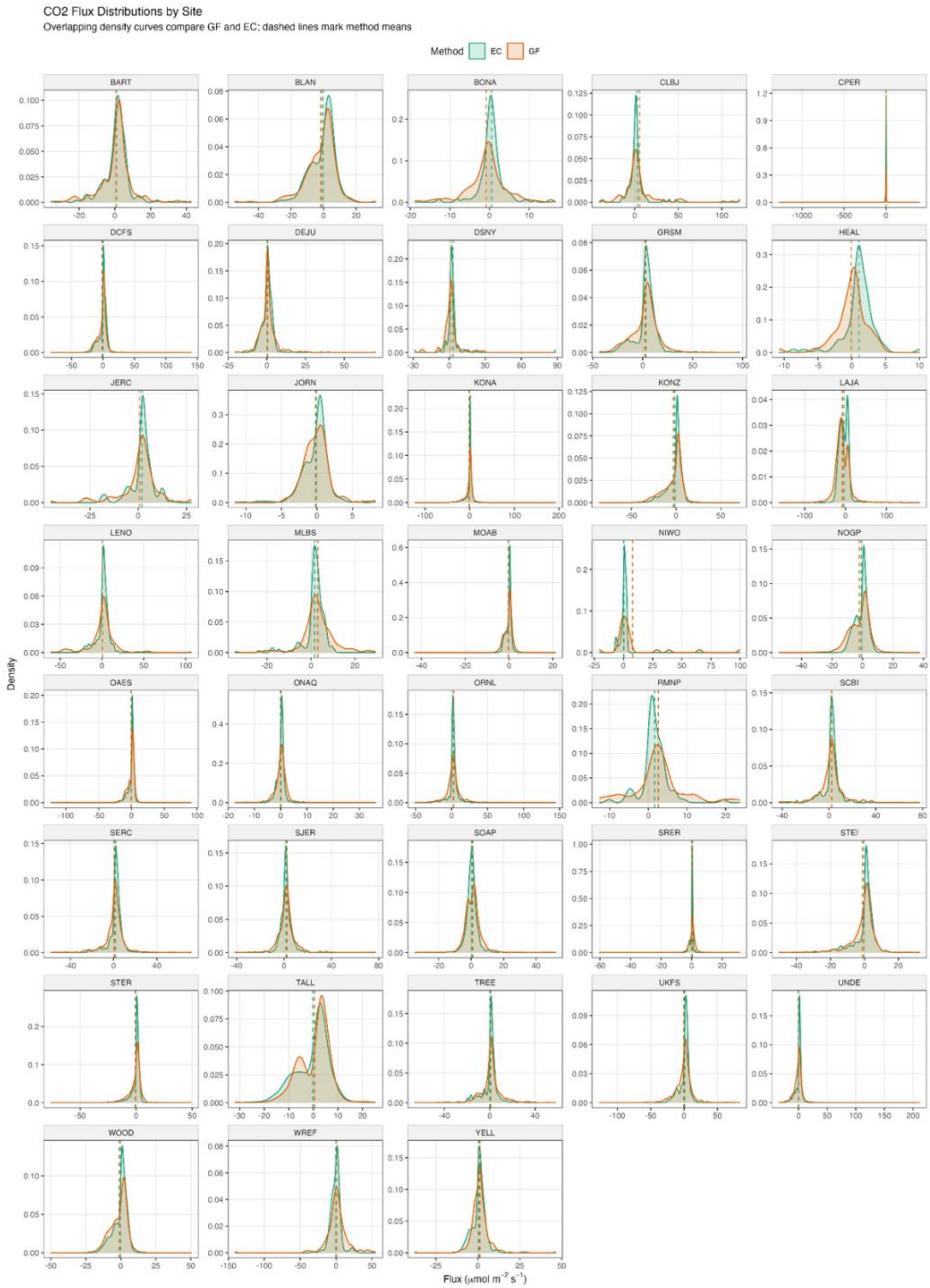


Figure A13. Overlapping density of half-hourly ensemble gradient (GF) and eddy-covariance (EC) fluxes for CO_2.

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

Open Access Earth System Science Data Discussions

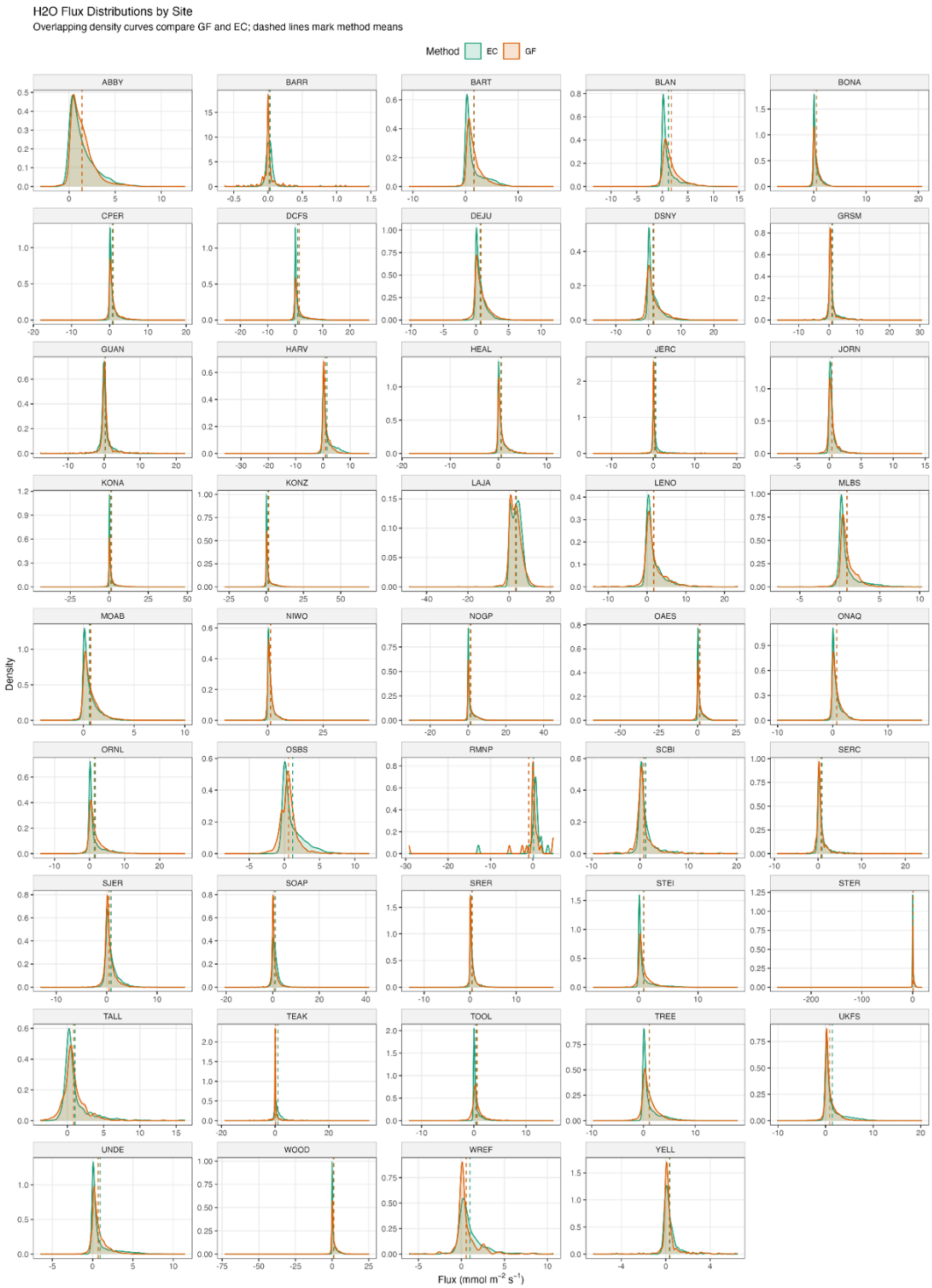
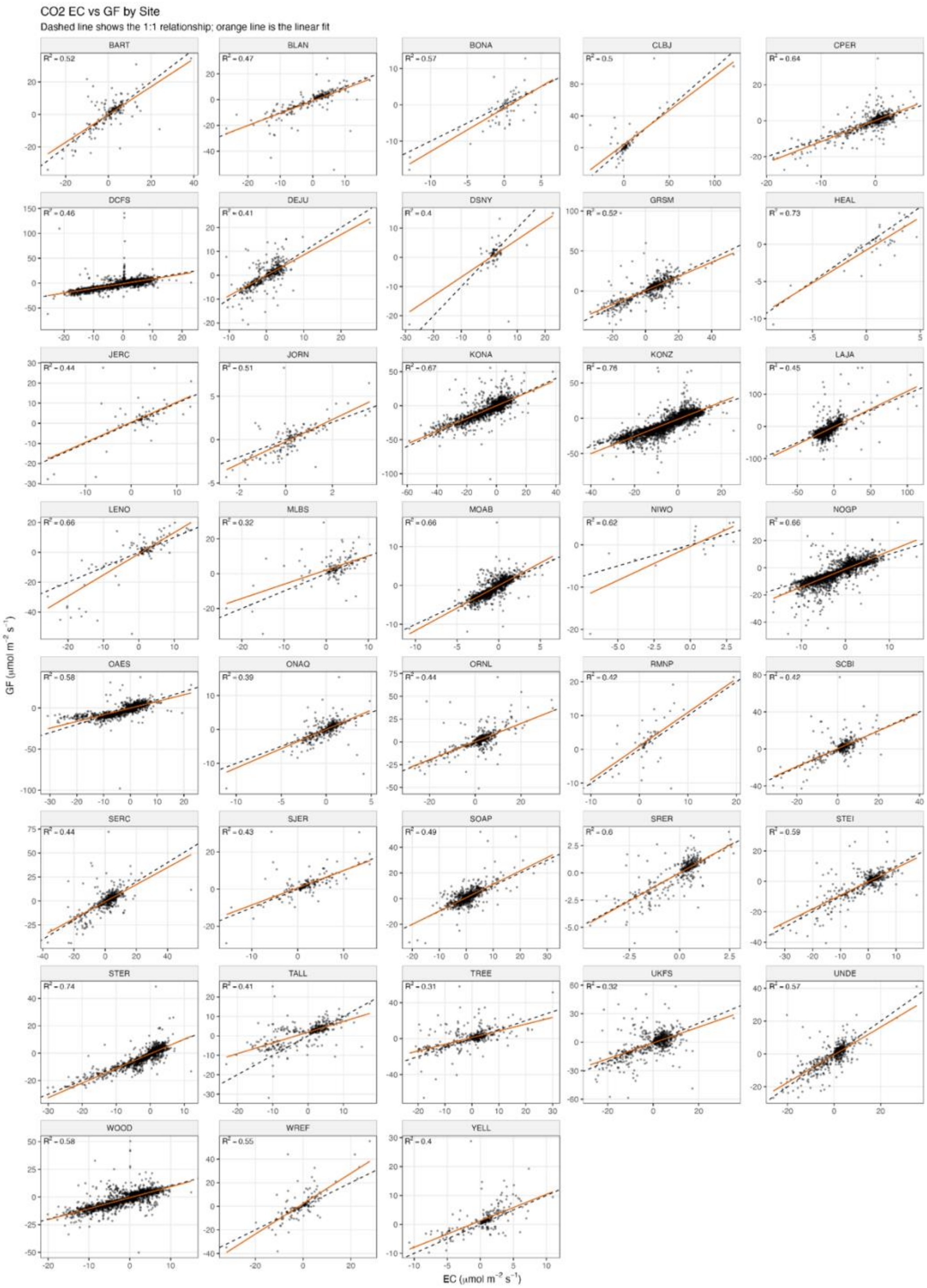


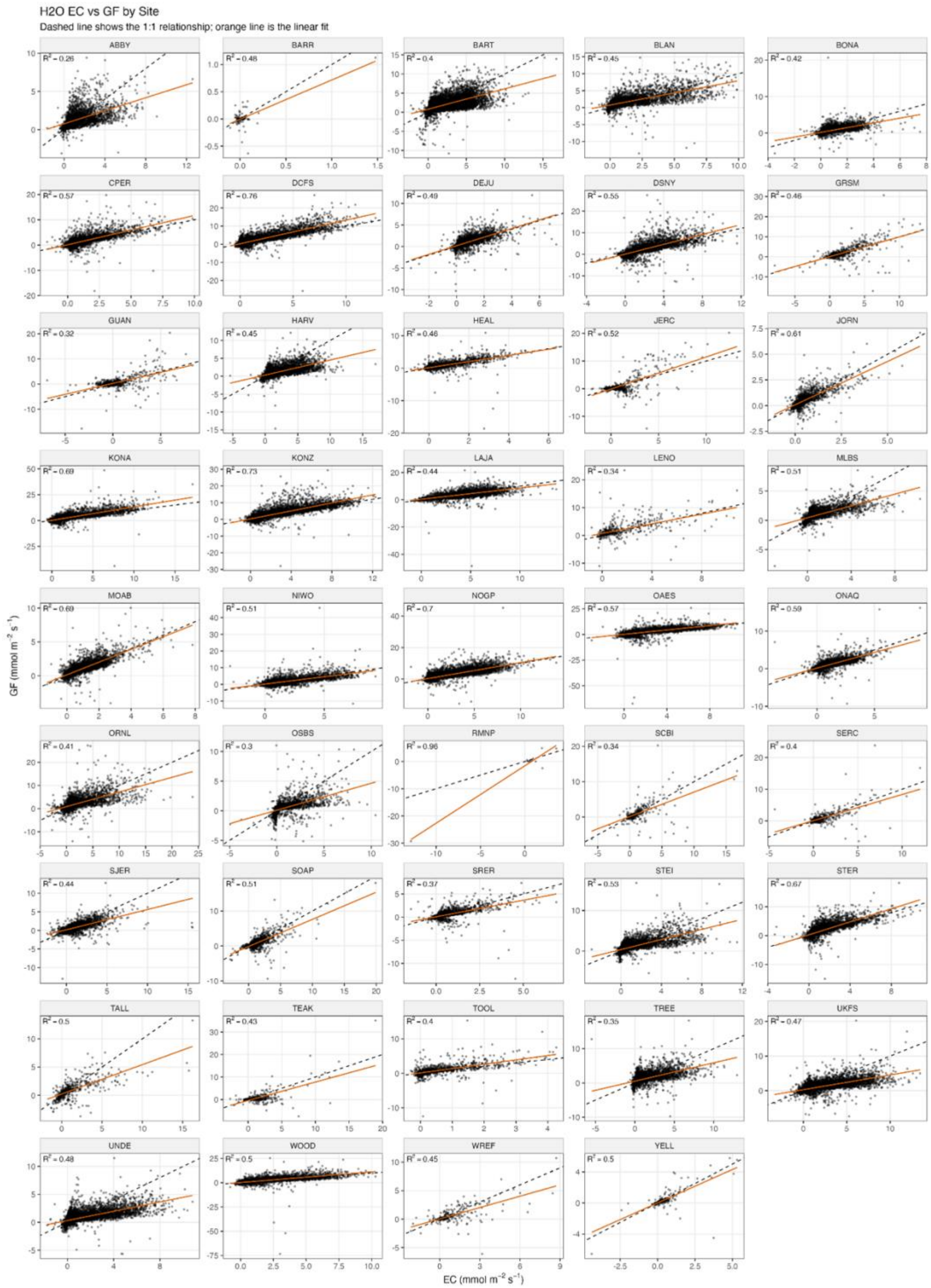
Figure A14. Overlapping density of half-hourly ensemble gradient (GF) and eddy-covariance (EC) fluxes for H_2O.

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026



655 **Figure A15. Half-hourly eddy covariance (EC) versus ensemble gradient (GF) fluxes for CO_2. The dashed line is the one-to-one line, and the orange line is the linear fit.**

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026



660 **Figure A16. Half-hourly eddy covariance (EC) versus ensemble gradient (GF) fluxes for CO_2. The dashed line is the one-to-one line, and the orange line is the linear fit.**

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

Open Access Earth System Science Data Discussions

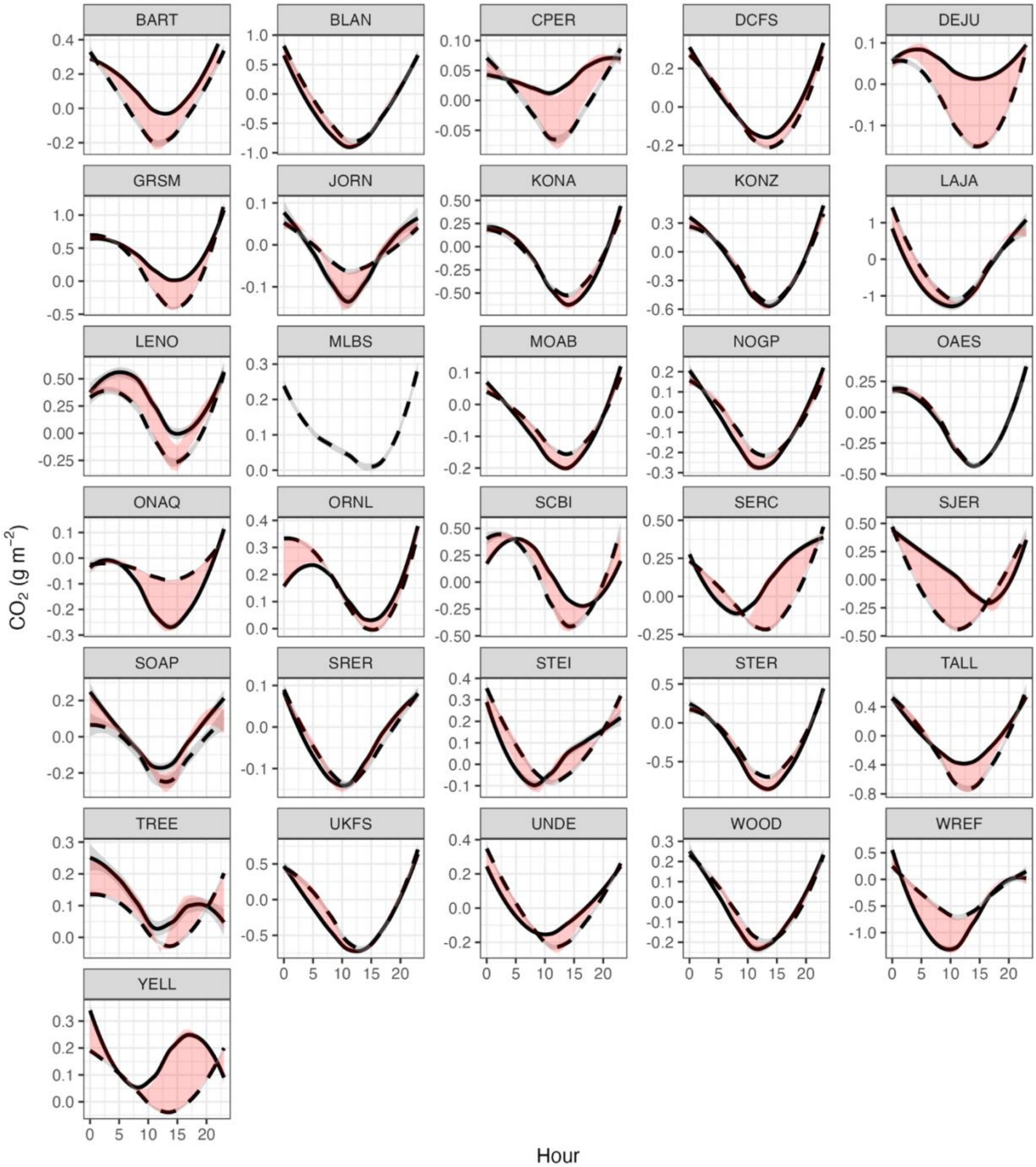
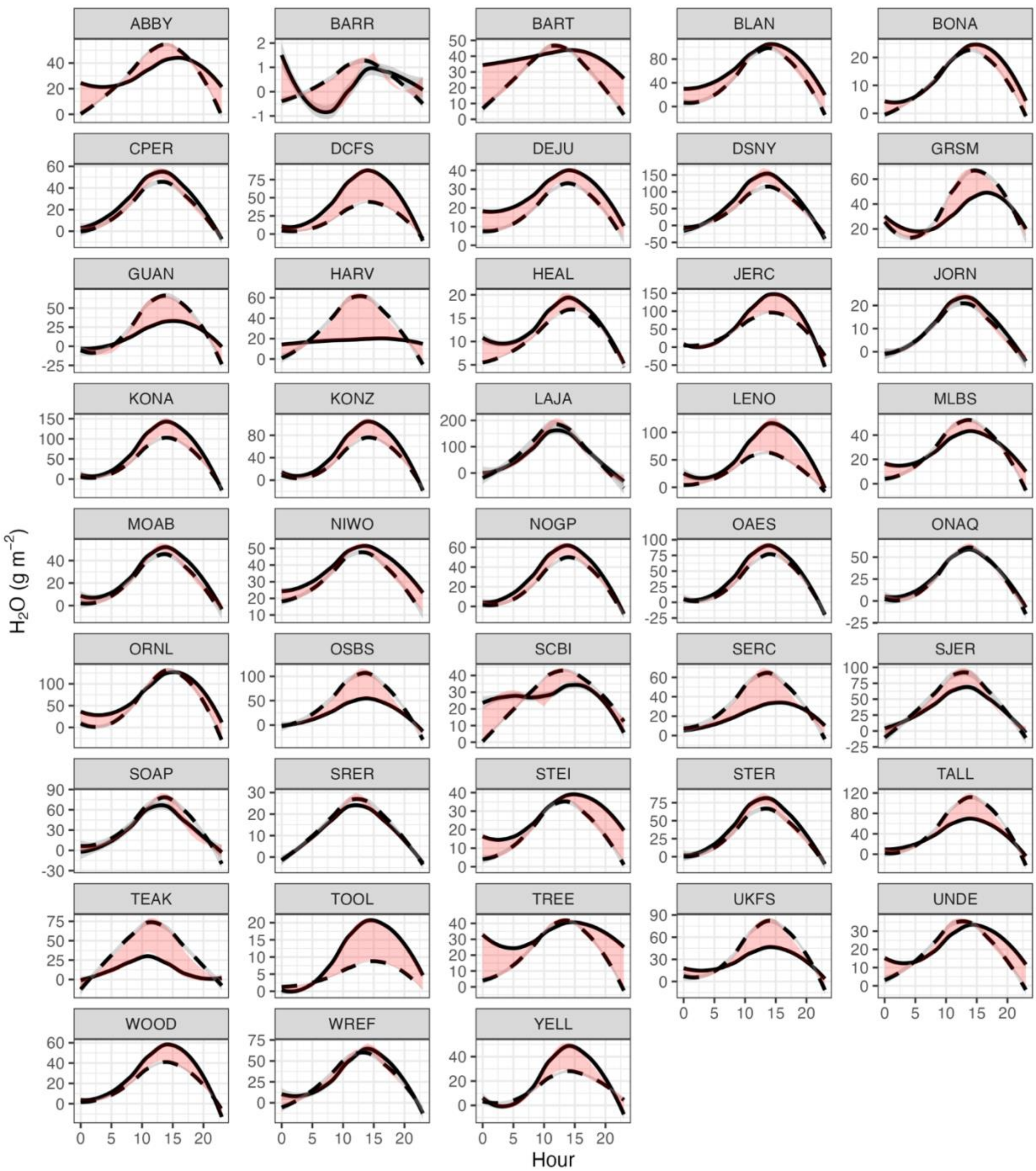


Figure A17. Ensemble CO_2 gradient (GF; solid) and eddy covariance (EC; dashed) diel curves for NEON sites in the summer.



665 **Figure A18. Ensemble H_2O gradient (GF; solid) and eddy covariance (EC; dashed) diel curves for NEON sites in the summer.**

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

Open Access Earth System Science Data Discussions

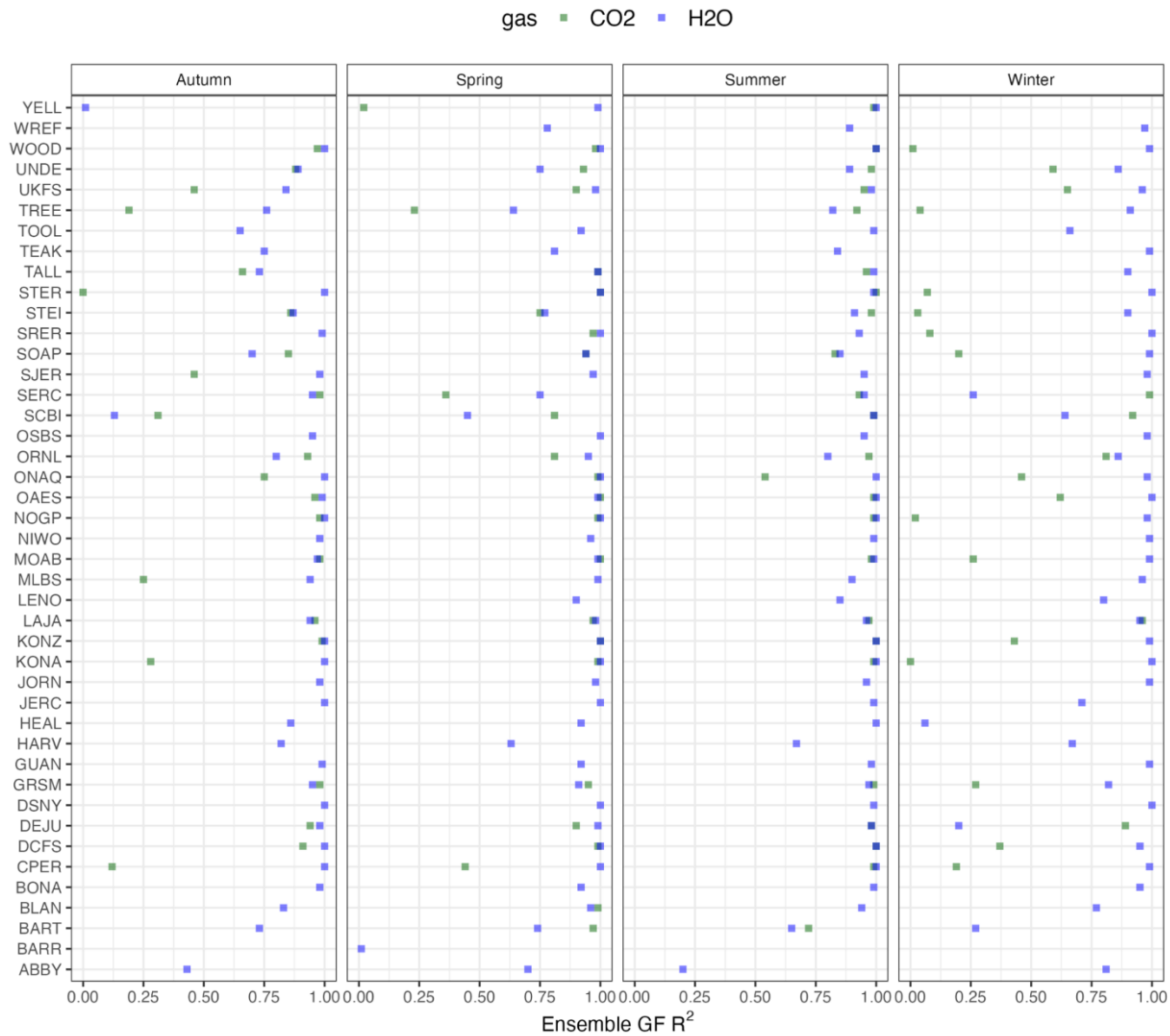


Figure A19. Evaluation of diel curves for CO_2 and H_2O by comparing hourly diels by season for ensemble gradient-fluxes (GF) and eddy covariance (EC).

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

Table A2. The range in total daily ensemble GF and EC based on seasonal diel curves for CO_2 (g CO_2 m^{-2} day^{-1}) and H_2O (g H_2O m^{-2} day^{-1}).

Site	Season	GF H_2O (g H_2O m^{-2} day^{-1})	GF CO_2 (g CO_2 m^{-2} day^{-1})	EC H_2O (g H_2O m^{-2} day^{-1})	EC CO_2 (g CO_2 m^{-2} day^{-1})
ABBY	Winter, Spring, Summer, Autumn	10.17 to 83.16	-	-20.53 to 128.58	-
BARR	Spring	-0.84 to 1.62	-	-0.37 to 1.42	-
BART	Winter, Spring, Summer, Autumn	13.65 to 119.08	-1.14 to 0.80	-9.36 to 173.44	-0.81 to 1.07
BLAN	Winter, Spring, Summer, Autumn	8.62 to 139.45	-1.17 to 0.76	-20.57 to 168.15	-1.07 to 0.78
BONA	Winter, Spring, Summer, Autumn	-2.70 to 51.60	-	-7.89 to 67.79	-
CPER	Winter, Spring, Summer, Autumn	-8.94 to 141.40	-0.57 to 0.27	-11.20 to 100.28	-0.45 to 0.27
DCFS	Winter, Spring, Summer, Autumn	-28.35 to 261.00	-1.16 to 0.52	-24.80 to 185.48	-1.03 to 0.73
DEJU	Winter, Spring, Summer, Autumn	2.00 to 74.94	-0.36 to 0.27	3.47 to 74.09	-0.38 to 0.35
DSNY	Winter, Spring, Summer, Autumn	-39.99 to 216.18	NA	-18.85 to 180.61	NA
GRSM	Winter, Spring, Summer, Autumn	-21.05 to 190.20	-1.10 to 1.39	-15.27 to 182.56	-1.38 to 1.42
GUAN	Winter, Spring, Summer, Autumn	-48.77 to 174.79	-	-24.66 to 100.78	-
HARV	Winter, Spring, Summer, Autumn	1.54 to 93.28	-	-7.04 to 190.71	-
HEAL	Winter, Spring, Summer, Autumn	-4.20 to 78.16	-	-2.36 to 70.59	-
JERC	Winter, Spring, Summer, Autumn	-51.20 to 150.79	-	-22.71 to 96.94	-
JORN	Winter, Spring, Summer, Autumn	-6.93 to 45.50	-0.16 to 0.05	-5.29 to 38.27	-0.07 to 0.04
KONA	Winter, Spring, Summer, Autumn	-36.36 to 260.92	-1.52 to 0.75	-30.83 to 185.10	-1.42 to 0.70
KONZ	Winter, Spring, Summer, Autumn	-31.38 to 257.90	-1.83 to 0.97	-29.95 to 191.93	-1.32 to 0.83
LAJA	Winter, Spring, Summer, Autumn	-11.62 to 204.04	-1.64 to 7.59	-34.12 to 209.82	-1.28 to 4.01
LENO	Winter, Spring, Summer, Autumn	-15.68 to 207.17	-2.55 to 0.78	-23.77 to 204.80	-1.59 to 0.73
MLBS	Winter, Spring, Summer, Autumn	9.86 to 88.68	-0.91 to 1.00	-22.23 to 142.97	-0.95 to 1.02
MOAB	Winter, Spring, Summer, Autumn	-1.92 to 65.99	-0.21 to 0.12	-8.32 to 62.72	-0.17 to 0.09
NIWO	Winter, Spring, Summer, Autumn	-3.59 to 158.59	-	-1.77 to 142.94	-
NOGP	Winter, Spring, Summer, Autumn	-21.81 to 189.36	-0.82 to 0.48	-26.72 to 168.27	-0.54 to 0.46
OAES	Winter, Spring, Summer, Autumn	-25.65 to 187.43	-0.76 to 0.48	-25.81 to 161.98	-0.83 to 0.49
ONAQ	Winter, Spring, Summer, Autumn	-4.45 to 62.82	-0.29 to 0.23	-10.85 to 67.04	-0.09 to 0.07
ORNL	Winter, Spring, Summer, Autumn	11.59 to 180.79	-0.90 to 0.76	-38.55 to 254.99	-0.96 to 0.44
OSBS	Winter, Spring, Summer, Autumn	-15.49 to 74.98	-	-19.34 to 147.39	-
SCBI	Winter, Spring, Summer, Autumn	-19.07 to 175.86	-1.13 to 0.81	-36.22 to 189.39	-0.96 to 0.63
SERC	Winter, Spring, Summer, Autumn	-14.60 to 98.82	-1.00 to 0.84	-16.41 to 160.42	-0.92 to 0.78
SJER	Winter, Spring, Summer, Autumn	-2.84 to 73.53	-0.28 to 0.45	-8.61 to 99.21	-0.45 to 0.44
SOAP	Winter, Spring, Summer, Autumn	-16.40 to 78.36	-0.20 to 0.37	-20.82 to 85.02	-0.31 to 0.20
SRER	Winter, Spring, Summer, Autumn	-1.35 to 40.84	-0.16 to 0.07	-1.77 to 56.80	-0.15 to 0.07
STEI	Winter, Spring, Summer, Autumn	1.64 to 103.04	-1.09 to 0.62	-21.95 to 147.42	-1.28 to 0.73
STER	Winter, Spring, Summer, Autumn	-8.83 to 147.76	-0.90 to 0.40	-5.18 to 92.50	-0.74 to 0.36
TALL	Winter, Spring, Summer, Autumn	-22.46 to 137.78	-0.62 to 0.66	-44.87 to 217.84	-1.01 to 0.67
TEAK	Winter, Spring, Summer, Autumn	-87.00 to 248.79	-	-37.05 to 156.75	-
TOOL	Winter, Spring, Summer, Autumn	-3.89 to 84.22	-	-3.38 to 54.63	-
TREE	Winter, Spring, Summer, Autumn	-0.37 to 100.69	-0.82 to 0.63	-15.64 to 141.73	-1.20 to 0.62
UKFS	Winter, Spring, Summer, Autumn	2.79 to 101.44	-1.35 to 0.60	-19.25 to 213.79	-1.24 to 0.77
UNDE	Winter, Spring, Summer, Autumn	0.45 to 68.77	-0.80 to 0.52	-12.12 to 143.69	-0.99 to 0.52
WOOD	Winter, Spring, Summer, Autumn	-16.42 to 172.97	-0.76 to 0.56	-15.83 to 138.38	-0.73 to 0.55
WREF	Winter, Spring, Summer, Autumn	-8.71 to 103.59	-1.38 to 1.26	-24.04 to 119.80	-0.75 to 0.43
YELL	Winter, Spring, Summer, Autumn	-4.08 to 78.98	-0.33 to 0.73	-4.17 to 97.51	-0.53 to 0.43

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

Code and data availability

The multi-site gradient-flux benchmark dataset is publicly archived via EDI:

Malone, S.L., A.R. Desai, J.D. Gewirtzman, C. Rey-Sanchez, S.A. Jurado, K.B. Delwiche, A. Chen, D.E. Reed, C.R. Florian, C.S. Sturtevant, J.H. Matthes, and R. Commane. 2026. Benchmark dataset for comparing eddy covariance measurements and gradient flux calculations across 47 National Ecological Observatory Network (NEON) flux towers, 2021-2024 ver 1. Environmental Data Initiative. (Last Access 2026-05-29). * Data is currently available for review here under the license BY 4.0:
https://portal.edirepository.org/nis/mapbrowse?scope=edi&identifier=2369&revision=1&key= LHRQgDvJXjj7acnumjrl-j3mEQ

Files include EC and GF fluxes by method and sampling height pair, along with corresponding micrometeorological data. Site-level canopy complexity from NEON AOP and metadata for tower geometry and sensor heights. The archive adheres to FAIR data principles by providing findable, DOI-based access, interoperable formats, reusable metadata, and transparent version control. Versioned DOI will be provided upon acceptance. Future releases may incorporate CH_4, additional NEON years, and additional benchmark methods.

Table 5. Benchmarking dataset contents.

Component	Coverage
Sites	47 terrestrial ecosystems
Ecosystem Types	5
Canopy height range	0.15–53 m
Sampling Height Pairs	+445
Canopy Levels	AA, AW
Datasets	EC, GF, storage fluxes, micrometerology, canopy complexity
Gases	CO_2, H_2O

Data download and flux calculation code is written in R and archived on GitHub:
Malone et al., (2026). lter/lterwg-flux-gradient: NEON Tower Gradient Flux Calculation (Version Manuscript1). Zenodo. https://doi.org/10.5281/zenodo.20597008

Malone et al., (2026). lter/lterwg-flux-gradient-eval: GRADIENT FLUX EVALUATION WORKFLOW (Version MANUSCRIPT1). Zenodo. https://doi.org/10.5281/zenodo.20597027

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

Earth System Science Data
Open Access Discussions

Author contributions

SLM, JM, CRS, KD, CF, CS, DR, SJ, RC, JG, and AD contributed to the project's conceptualization and investigation through participation in the invited workshop and monthly meetings. SLM provided project administration and supervision. CRS led the flux calculations. SLM, SJ, JM, KD, JG, CS, AC, and CRS performed the data analysis and project execution. CS and CF provided resources and data curation, contributing critical site-specific data and expertise to the synthesis. All authors contributed to the review and editing of the manuscript.

Acknowledgements

This work was conducted as a part of The Flux Gradient Project Synthesis Group funded by the National Science Foundation under grant DEB#1929393, through the Long Term Ecological Research Network Office (LNO), National Center for Ecological Analysis and Synthesis (NCEAS), University of California, Santa Barbara. The National Ecological Observatory Network is a program sponsored by the U.S. National Science Foundation and operated under cooperative agreement by Battelle. Data used in this research were obtained through NEON Research Support Services. We also acknowledge the use of AmeriFlux data resources provided by the U.S. Department of Energy's Office of Science. Support for this work was also provided by the Yale Center for Natural Carbon Capture and the Yale School of the Environment. MU was supported by the Arctic Challenge for Sustainability III (ArCS III; JPMXD1720251001). Authors recognize Nicholas J. Lyon, from the LTER network office for his guidance on data science and code structure.

References

Baldocchi, D.: Measuring fluxes of trace gases and energy between ecosystems and the atmosphere - the state and future of the eddy covariance method, Glob. Chang. Biol., 20, 3600–3609, https://doi.org/10.1111/gcb.12649, 2014.

Baldocchi, D. D.: Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future, Glob. Chang. Biol., 9, 479–492, 2003.

Baldocchi, D. D. and Meyers, T. P.: Trace gas exchange above the floor of a deciduous forest: 1. Evaporation and CO_2 efflux, J. Geophys. Res., 96, 7271–7285, https://doi.org/10.1029/91jd00269, 1991.

Bolinius, D. J., Jahnke, A., and MacLeod, M.: Comparison of eddy covariance and modified Bowen ratio methods for measuring gas fluxes and implications for measuring fluxes of persistent organic pollutants, Atmos. Chem. Phys., 16, 5315–5322, https://doi.org/10.5194/acp-16-5315-2016, 2016.

Brock, F. V.: A Nonlinear Filter to Remove Impulse Noise from Meteorological Data, J. Atmos. Ocean. Technol., 3, 51–58, https://doi.org/10.1175/1520-0426(1986)003<0051:ANFTRI>2.0.CO;2, 1986.

Brown, S. E., Wagner-Riddle, C., and Conrad, B.: Low-power flux gradient measurements for quantifying the impact of agricultural management on nitrous oxide emissions, Agric. For. Meteorol., 353, 110027, https://doi.org/10.1016/j.agrformet.2024.110027, 2024.

Cellier, P. and Brunet, Y.: Flux-gradient relationships above tall plant canopies, Agric. For. Meteorol., 58, 93–117, https://doi.org/10.1016/0168-1923(92)90113-I, 1992.

Commane, R., Meredith, L. K., Baker, I. T., Berry, J. A., Munger, J. W., Montzka, S. A., Templer, P. H., Juice, S. M., Zahniser, M. S., and Wofsy, S. C.: Seasonal fluxes of carbonyl sulfide in a midlatitude forest, Proc. Natl. Acad. Sci. U. S. A., 112, 14162–14167, https://doi.org/10.1073/pnas.1504131112, 2015.

Dyer, A. J. and Hicks, B. B.: Flux-gradient relationships in the constant flux layer, Quart. J. Roy. Meteor. Soc., 96, 715–721, https://doi.org/10.1002/qj.49709641012, 1970.

Edwards, G. C., Thurtell, G. W., Kidd, G. E., Dias, G. M., and Wagner-Riddle, C.: A diode laser based gas monitor suitable for measurement of trace gas exchange using micrometeorological techniques, Agric. For. Meteorol., 115, 71–89, https://doi.org/10.1016/s0168-1923(02)00166-1, 2003.

Foken, T. and Leclerc, M. Y.: Methods and limitations in validation of footprint models, Agric. For. Meteorol., 127, 223–234, https://doi.org/10.1016/j.agrformet.2004.07.015, 2004.

Garratt, J. R.: Flux profile relations above tall vegetation, Q. J. R. Meteorol. Soc., 104, 199–211, https://doi.org/10.1002/qj.49710443915, 1978.

Genuer, R., Poggi, J.-M., and Tuleau-Malot, C.: VSURF: an R package for variable selection using random forests, 2015.

Griffith, D. W. T., Leuning, R., Denmead, O. T., and Jamie, I. M.: Air–land exchanges of CO2, CH4 and N2O measured by FTIR spectrometry and micrometeorological techniques, Atmos. Environ. (1994), 36, 1833–1842, https://doi.org/10.1016/s1352-2310(02)00139-5, 2002.

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and Wang, X.: The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions, Geosci. Model Dev., 5, 1471–1492, https://doi.org/10.5194/gmd-5-1471-2012, 2012.

Gu, L., Falge, E. M., Boden, T., Baldocchi, D. D., Black, T. A., Saleska, S. R., Suni, T., Verma, S. B., Vesala, T., Wofsy, S. C., and Xu, L.: Objective threshold determination for nighttime eddy flux filtering, Agric. For. Meteorol., 128, 179–197, https://doi.org/10.1016/j.agrformet.2004.11.006, 2005.

Hall, B., Claiborn, C., and Baldocchi, D.: Measurement and modeling of the dry deposition of peroxides, Atmos. Environ. (1994), 33, 577–589, https://doi.org/10.1016/s1352-2310(98)00271-4, 1999.

Hicks, B. B. and Baldocchi, D. D.: Measurement of fluxes over land: Capabilities, origins, and remaining challenges, Boundary Layer Meteorol., 177, 365–394, https://doi.org/10.1007/s10546-020-00531-y, 2020.

Högström, U., Bergström, H., Smedman, A.-S., Halldin, S., and Lindroth, A.: Turbulent exchange above a pine forest, I: Fluxes and gradients, Boundary Layer Meteorol., 49, 197–217, https://doi.org/10.1007/bf00116411, 1989.

Kamp, J. N., Häni, C., Nyord, T., Feilberg, A., and Sørensen, L. L.: The aerodynamic gradient method: Implications of non-simultaneous measurements at alternating heights, Atmosphere (Basel), 11, 1067, https://doi.org/10.3390/atmos11101067, 2020.

Kao, R. H., Gibson, C. M., Rachel E. Gallery, Meier, C. L., Barnett, D. T., Docherty, K. M., Blevins, K. K., Travers, P. D., Azuaje, E., Springer, Y. P., Thibault, K. M., McKenzie, V. J., Keller, M., Alves, L. F., Hinckley, E.-L. S., Parnell, J., and Schimel, D.: NEON terrestrial field observations: designing continental-scale, standardized sampling, Ecosphere, 3, art115, https://doi.org/10.1890/es12-00196.1, 2012.

Karl, T., Guenther, A., Spirig, C., Hansel, A., and Fall, R.: Seasonal variation of biogenic VOC emissions above a mixed hardwood forest in northern Michigan: BIOGENIC VOC EMISSIONS ABOVE MIXED HARDWOOD FOREST, Geophys. Res. Lett., 30, https://doi.org/10.1029/2003gl018432, 2003.

Kenion, H. C., Davis, K. J., Miles, N. L., Monteiro, V. C., Richardson, S. J., and Horne, J. P.: Estimation of urban greenhouse gas fluxes from mole fraction measurements using Monin–Obukhov similarity theory, J. Atmos. Ocean. Technol., 41, 833–846, https://doi.org/10.1175/jtech-d-23-0164.1, 2024.

Landis, J. R. and Koch, G. G.: The measurement of observer agreement for categorical data, Biometrics, 33, 159–174, https://doi.org/10.2307/2529310, 1977.

LaRue, E. A., Fahey, R., Fuson, T. L., Foster, J. R., Matthes, J. H., Krause, K., and Hardiman, B. S.: Evaluating the sensitivity of forest structural diversity characterization to LiDAR point density, Ecosphere, 13, e4209, https://doi.org/10.1002/ecs2.4209, 2022.

Laubach, J., Barthel, M., Fraser, A., Hunt, J. E., and Griffith, D. W. T.: Combining two complementary micrometeorological methods to measure CH_4 and N_2O fluxes over pasture, Biogeosciences, 13, 1309–1327, https://doi.org/10.5194/bg-13-1309-2016, 2016.

Lee, E.-H., Lee, E., Park, R., Kwon, Y. C., and Hong, S.-Y.: Impact of turbulent mixing in the stratocumulus-topped boundary layer on numerical weather prediction, Asia Pac. J. Atmos. Sci., 54, 371–384, https://doi.org/10.1007/s13143-018-0024-0, 2018.

Liaw, A. and Wiener, M.: Classification and regression by randomForest, R news, 2, 18–22, 2002.

Lin, L. I.: A concordance correlation coefficient to evaluate reproducibility, Biometrics, 45, 255–268, https://doi.org/10.2307/2532051, 1989.

Maier, M. and Schack-Kirchner, H.: Using the gradient method to determine soil gas flux: A review, Agric. For. Meteorol., 192-193, 78–95, https://doi.org/10.1016/j.agrformet.2014.03.006, 2014.

Malone, S. L., Oh, Y., Arndt, K. A., Burba, G., Commane, R., Contosta, A. R., Goodrich, J. P., Loescher, H. W., Starr, G., and Varner, R. K.: Gaps in network infrastructure limit our understanding of biogenic methane emissions for the United States, Biogeosciences, 19, 2507–2522, 2022.

Malone, S.L., A.R. Desai, J.D. Gewirtzman, C. Rey-Sanchez, S.A. Jurado, K.B. Delwiche, A. Chen, D.E. Reed, C.R. Florian, C.S. Sturtevant, J.H. Matthes, and R. Commane. 2026. Benchmark dataset for comparing eddy covariance 670 measurements and gradient flux calculations across 47 National Ecological Observatory Network (NEON) flux towers, 2021-2024 ver 1. Environmental Data Initiative. https://doi.org/DOI_PLACE_HOLDER (Accessed 2026-05-29).

Malone, S.L., A.R. Desai, J.D. Gewirtzman, C. Rey-Sanchez, S.A. Jurado, K.B. Delwiche, A. Chen, D.E. Reed, C.R. Florian, C.S. Sturtevant, J.H. Matthes, and R. Commane. (2026). lter/lterwg-flux-gradient: NEON Tower Gradient Flux Calculation (Version Manuscript1). Zenodo. https://doi.org/10.5281/zenodo.20597008

Malone, S.L., A.R. Desai, J.D. Gewirtzman, C. Rey-Sanchez, S.A. Jurado, K.B. Delwiche, A. Chen, D.E. Reed, C.R. Florian, C.S. Sturtevant, J.H. Matthes, and R. Commane. (2026). lter/lterwg-flux-gradient-eval: GRADIENT FLUX EVALUATION WORKFLOW (Version MANUSCRIPT1). Zenodo. https://doi.org/10.5281/zenodo.20597027

Mayer, J.-C., Bargsten, A., Rummel, U., Meixner, F. X., and Foken, T.: Distributed Modified Bowen Ratio method for surface layer fluxes of reactive and non-reactive trace gases, Agric. For. Meteorol., 151, 655–668, https://doi.org/10.1016/j.agrformet.2010.10.001, 2011.

Meredith, L. K., Commane, R., Munger, J. W., Dunn, A., Tang, J., Wofsy, S. C., and Prinn, R. G.: Ecosystem fluxes of hydrogen: a comparison of flux-gradient methods, Atmospheric Measurement Techniques, 7, 2787–2805, 2014.

Metzger, S., Burba, G., Burns, S. P., Blanken, P. D., Li, J., Luo, H., and Zulueta, R. C.: Optimization of an enclosed gas analyzer sampling system for measuring eddy covariance fluxes of H 2 O and CO 2, Atmospheric Measurement Techniques, 9, 1341–1359, 2016.

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, Geosci. Model Dev., 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., and Zulueta R: 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, 2019.

Meyers, T. P. and Baldocchi, D. D.: Current micrometeorological flux methodologies with applications in agriculture, in: Agronomy Monographs, American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, Madison, WI, USA, 381–396, https://doi.org/10.2134/agronmonogr47.c16, 2015.

Meyers, T. P., Hall, M. E., Lindberg, S. E., and Kim, K.: Use of the modified bowen-ratio technique to measure fluxes of trace gases, Atmos. Environ., 30, 3321–3329, https://doi.org/10.1016/1352-2310(96)00082-9, 1996.

Miyata, A., Leuning, R., Denmead, O. T., Kim, J., and Harazono, Y.: Carbon dioxide and methane fluxes from an intermittently flooded paddy field, Agric. For. Meteorol., 102, 287–303, https://doi.org/10.1016/s0168-1923(00)00092-7, 2000.

Mölder, M., Grelle, A., Lindroth, A., and Halldin, S.: Flux-profile relationships over a boreal forest — roughness sublayer corrections, Agric. For. Meteorol., 98-99, 645–658, https://doi.org/10.1016/s0168-1923(99)00131-8, 1999.

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

Moncrieff, J. B., Massheder, J. M., de Bruin, H., Elbers, J., Friborg, T., Heusinkveld, B., Kabat, P., Scott, S., Soegaard, H., and Verhoef, A.: A system to measure surface fluxes of momentum, sensible heat, water vapour and carbon dioxide, J. Hydrol., 188-189, 589–611, https://doi.org/10.1016/S0022-1694(96)03194-0, 1997.

Monteith, J. and Unsworth, M.: Principles of environmental physics: plants, animals, and the atmosphere, 2013.

Nelson, A. J., Lichiheb, N., Koloutsou-Vakakis, S., Rood, M. J., Heuer, M., Myles, L., Joo, E., Miller, J., and Bernacchi, C.: Ammonia flux measurements above a corn canopy using relaxed eddy accumulation and a flux gradient system, Agric. For. Meteorol., 264, 104–113, https://doi.org/10.1016/j.agrformet.2018.10.003, 2019.

NEON: 2D wind speed and direction (DP1.00001.001), 2025a.

NEON: Bundled data products - eddy covariance (DP4.00200.001), 2025b.

NEON: Ecosystem structure (DP3.30015.001), 2025c.

NEON: Elevation - LiDAR (DP3.30024.001), 2025d.

NEON: LAI - spectrometer - mosaic (DP3.30012.001), 2025e.

NEON: Photosynthetically active radiation (PAR) (DP1.00024.001), 2025f.

NEON: Relative humidity (DP1.00098.001), 2025g.

NEON: Vegetation indices - spectrometer - mosaic (DP3.30026.001), 2025h.

Obrist, D., Roy, E. M., Harrison, J. L., Kwong, C. F., Munger, J. W., Moosmüller, H., Romero, C. D., Sun, S., Zhou, J., and Commane, R.: Previously unaccounted atmospheric mercury deposition in a midlatitude deciduous forest, Proc. Natl. Acad. Sci. U. S. A., 118, e2105477118, https://doi.org/10.1073/pnas.2105477118, 2021.

Okoye, K. and Hosseini, S.: Mann–Whitney U test and kruskal–Wallis H test statistics in R, in: R Programming, Springer Nature Singapore, Singapore, 225–246, https://doi.org/10.1007/978-981-97-3385-9_11, 2024.

Ostertagova, E., Ostertag, O., and Kováč, J.: Methodology and application of the Kruskal-Wallis test, Applied mechanics and materials, 611, 115–120, 2014.

Raupach, M. R.: Anomalies in flux-gradient relationships over forest, Boundary Layer Meteorol., 16, 467–486, https://doi.org/10.1007/bf03163564, 1979.

Raupach, M. R. and Legg, B. J.: The uses and limitations of flux-gradient relationships in micrometeorology, Agric. Water Manag., 8, 119–131, https://doi.org/10.1016/0378-3774(84)90049-0, 1984.

Raupach, M. R. and Thom, A. S.: Turbulence in and above Plant Canopies, Annu. Rev. Fluid Mech., 13, 97–129, https://doi.org/10.1146/annurev.fl.13.010181.000525, 1981.

R Core Team: R: A Language and Environment for Statistical Computing, 2021.

Rebmann, C., Kolle, O., Heinesch, B., Queck, R., Ibrom, A., and Aubinet, M.: Data Acquisition and Flux Calculations, in: Eddy Covariance, Springer Netherlands, Dordrecht, 59–83, https://doi.org/10.1007/978-94-007-2351-1_3, 2012.

Reed, D. E., Chu, H., Peter, B. G., Chen, J., Abraha, M., Amiro, B., Anderson, R. G., Arain, M. A., Arruda, P., Barron-Gafford, G. A., Bernacchi, C., Beverly, D. P., Biraud, S. C., Black, T. A., Blanken, P. D., Bohrer, G., Bowler, R., Bowling, D. R., Bret-Harte, M. S., Bretfeld, M., Brunsell, N. A., Bullock, S. H., Celis, G., Chen, X., Classen, A. T., Cook, D. R., Cueva, A.,

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

Dalmagro, H. J., Davis, K., Desai, A., Duff, A. J., Dunn, A. L., Durden, D., Edgar, C. W., Euskirchen, E., Bracho, R., Ewers, B., Flanagan, L. B., Florian, C., Foord, V., Forbrich, I., Forsythe, B. R., Frank, J., Garatuza-Payan, J., Goslee, S., Gough, C., Green, M., Griffis, T., Helbig, M., Hill, A. C., Hinkle, R., Horne, J., Humphreys, E., Ikawa, H., Iwahana, G., Jassal, R., Johnson, B., Johnson, M., Kannenberg, S. A., Kelsey, E., King, J., Knowles, J. F., Knox, S., Kobayashi, H., Kolb, T., Kolka, R., Krauss, K. W., Kutzbach, L., Lamb, B., Law, B., Lee, S.-C., Lee, X., Liu, H., Loescher, H. W., Malone, S. L., Matamala, R., Mauritz, M., Metzger, S., Meyer, G., Mitra, B., Munger, J. W., Nesic, Z., Noormets, A., O'Halloran, T. L., O'Keeffe, P. T., Oberbauer, S. F., Oechel, W., Oikawa, P., Olivas, P. C., Ouimette, A., Pastorello, G., Perez-Quezada, J. F., Phillips, C., Posse, G., Qu, B., Quinton, W. L., Reba, M. L., Richardson, A. D., Picasso, V., et al.: Network of networks: Time series clustering of AmeriFlux sites, Agric. For. Meteorol., 372, 110686, https://doi.org/10.1016/j.agrformet.2025.110686, 2025.

Schotanus, P., Nieuwstadt, F. T. M., and De Bruin, H. A. R.: Temperature measurement with a sonic anemometer and its application to heat and moisture fluxes, Boundary Layer Meteorol., 26, 81–93, https://doi.org/10.1007/bf00164332, 1983.

Simpson, I. J., Edwards, G. C., Thurtell, G. W., den Hartog, G., Neumann, H. H., and Staebler, R. M.: Micrometeorological measurements of methane and nitrous oxide exchange above a boreal aspen forest, J. Geophys. Res., 102, 29331–29341, https://doi.org/10.1029/97jd03181, 1997.

Simpson, I. J., Thurtell, G. W., Neumann, H. H., Den Hartog, G., and Edwards, G. C.: The Validity of Similarity Theory in the Roughness Sublayer Above Forests, Bound.-Layer Meteorol., 87, 69–99, https://doi.org/10.1023/A:1000809902980, 1998.

Smith, P., Ashmore, M. R., Black, H. I. J., Burgess, P. J., Evans, C. D., Quine, T. A., Thomson, A. M., Hicks, K., and Orr, H. G.: REVIEW: The role of ecosystems and their management in regulating climate, and soil, water and air quality, J. Appl. Ecol., 50, 812–829, https://doi.org/10.1111/1365-2664.12016, 2013.

Spirig, C., Neftel, A., Ammann, C., Dommen, J., Grabmer, W., Thielmann, A., Schaub, A., Beauchamp, J., Wisthaler, A., and Hansel, A.: Eddy covariance flux measurements of biogenic VOCs during ECHO 2003 using proton transfer reaction mass spectrometry, Atmos. Chem. Phys., 5, 465–481, https://doi.org/10.5194/acp-5-465-2005, 2005.

Starkenburg, D., Metzger, S., Fochesatto, G. J., Alfieri, J. G., Gens, R., Prakash, A., and Cristóbal, J.: Assessment of Despiking Methods for Turbulence Data in Micrometeorology, https://doi.org/10.1175/jtech-d-15-0154.1, 2016.

Stewart, J. B. and Thom, A. S.: Energy budgets in pine forest, Q. J. R. Meteorol. Soc., 99, 154–170, https://doi.org/10.1002/qj.49709941913, 1973.

Stull, R. B.: An Introduction to Boundary Layer Meteorology Springer, 1988.

Stull, R. B.: An introduction to boundary layer meteorology, 13, 2012.

Sturtevant, C., DeRego, E., Metzger, S., Ayres, E., Allen, D., Burlingame, T., Catolico, N., Cawley, K., Csavina, J., Durden, D., Florian, C., Frost, S., Gaddie, R., Knapp, E., Laney, C., Lee, R., Lenz, D., Litt, G., Luo, H., Roberti, J., Slemmons, C., Styers, K., Tran, C., Vance, T., and SanClements, M.: A process approach to quality management doublesNEONsensor data quality, Methods Ecol. Evol., 13, 1849–1865, https://doi.org/10.1111/2041-210x.13943, 2022.

https://doi.org/10.5194/essd-2026-413
Preprint. Discussion started: 31 July 2026

Tanimoto, H., Kameyama, S., Iwata, T., Inomata, S., and Omori, Y.: Measurement of air-sea exchange of dimethyl sulfide and acetone by PTR-MS coupled with gradient flux technique, Environ. Sci. Technol., 48, 526–533, https://doi.org/10.1021/es4032562, 2014.

Wagner-Riddle, C., Thurtell, G. W., King, K. M., Kidd, G. E., and Beauchamp, E. G.: Nitrous oxide and carbon dioxide fluxes from a bare soil using a micrometeorological approach, J. Environ. Qual., 25, 898–907, https://doi.org/10.2134/jeq1996.00472425002500040035x, 1996.

Wagner-Riddle, C., Thurtell, G. W., and Edwards, G. C.: Trace Gas Concentration Measurements for Micrometeorological Flux Quantification, in: Agronomy Monographs, American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, Madison, WI, USA, 321–343, https://doi.org/10.2134/agronmonogr47.c14, 2005.

Wang, J., Choi, D. H., LaRue, E., Atkins, J. W., Foster, J. R., Matthes, J. H., Fahey, R. T., Fei, S., and Hardiman, B. S.: NEON-SD: A 30-m structural diversity product derived from the NEON Discrete-Return LiDAR Point Cloud, Sci. Data, 11, 1174, https://doi.org/10.1038/s41597-024-04018-0, 2024.

Welles, J. M. and McDermitt, D. K.: Measuring carbon dioxide in the atmosphere, in: Agronomy Monographs, American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, Madison, WI, USA, 287–320, https://doi.org/10.2134/agronmonogr47.c13, 2005.

Werner, C., Davis, K., Bakwin, P., Yi, C., Hurst, D., and Lock, L.: Regional-scale measurements of CH_4 exchange from a tall tower over a mixed temperate/boreal lowland and wetland forest, Glob. Chang. Biol., 9, 1251–1261, https://doi.org/10.1046/j.1365-2486.2003.00670.x, 2003.

Willcock, S., Hooftman, D. A. P., Blanchard, R., Dawson, T. P., Hickler, T., Lindeskog, M., Martinez-Lopez, J., Reyers, B., Watts, S. M., Eigenbrod, F., and Bullock, J. M.: Ensembles of ecosystem service models can improve accuracy and indicate uncertainty, Sci. Total Environ., 747, 141006, https://doi.org/10.1016/j.scitotenv.2020.141006, 2020.

Wu, Z. Y., Zhang, L., Wang, X. M., and Munger, J. W.: A modified micrometeorological gradient method for estimating O_3 dry depositions over a forest canopy, Atmos. Chem. Phys., 15, 7487–7496, https://doi.org/10.5194/acp-15-7487-2015, 2015.

Yuan, R., Kang, M., Park, S.-B., Hong, J., Lee, D., and Kim, J.: Expansion of the planar-fit method to estimate flux over complex terrain, Meteorol. Atmos. Phys., 110, 123–133, https://doi.org/10.1007/s00703-010-0113-9, 2011.

Zhao, J., Zhang, M., Xiao, W., Wang, W., Zhang, Z., Yu, Z., Xiao, Q., Cao, Z., Xu, J., Zhang, X., Liu, S., and Lee, X.: An evaluation of the flux-gradient and the eddy covariance method to measure CH4, CO2, and H2O fluxes from small ponds, Agric. For. Meteorol., 275, 255–264, https://doi.org/10.1016/j.agrformet.2019.05.032, 2019.