

Lawrence Berkeley National Laboratory

LBL Publications

Title

A tale of two towers: comparing NEON and AmeriFlux data streams at Bartlett Experimental Forest

Permalink

<https://escholarship.org/uc/item/5xb120jf>

Journal

Agricultural and Forest Meteorology, 378

ISSN

0168-1923

Authors

Liu, Yujie

Stoy, Paul

Chu, Housen

et al.

Publication Date

2026-03-01

DOI

10.1016/j.agrformet.2025.110939

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

A tale of two towers: comparing NEON and AmeriFlux data streams at Bartlett Experimental Forest

Contact information: Yujie Liu (Yujie.Liu_YL@outlook.com, yujie.liu@nau.edu)

Co-authors and affiliations

Yujie Liu (ORCID: 0000-0003-0335-6400): Northern Arizona University, Center for Ecosystem Science and Society (ECOSS), AZ 86004, USA.

Paul Stoy (ORCID: 0000-0002-6053-6232): Department of Biological Systems Engineering, University of Wisconsin – Madison, Madison, WI, USA.

Housen Chu (ORCID: 0000-0002-8131-4938): Climate & Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA.

Dave Y. Hollinger (ORCID: 0000-0002-4284-1575): U. S. Department of Agriculture, Forest Service, Northern Research Station, Durham, NH, USA.

Scott V. Ollinger (ORCID: 0000-0001-6226-1431): Earth Systems Research Center, University of New Hampshire, Durham, NH, USA.

Andrew P. Ouimette (ORCID: 0000-0003-3016-5730): U.S. Department of Agriculture, Forest Service, Northern Research Station, Durham, NH, USA.

David J. Durden (ORCID: 0000-0001-9572-8325): National Ecological Observatory Network, Battelle, Boulder, CO, USA.

Cove Sturtevant (ORCID: 0000-0002-0341-3228): National Ecological Observatory Network, Battelle, Boulder, CO, USA.

Ben Lucas (ORCID: 0000-0002-2021-3076): Northern Arizona University, Department of Mathematics and Statistics, AZ 86004, USA.

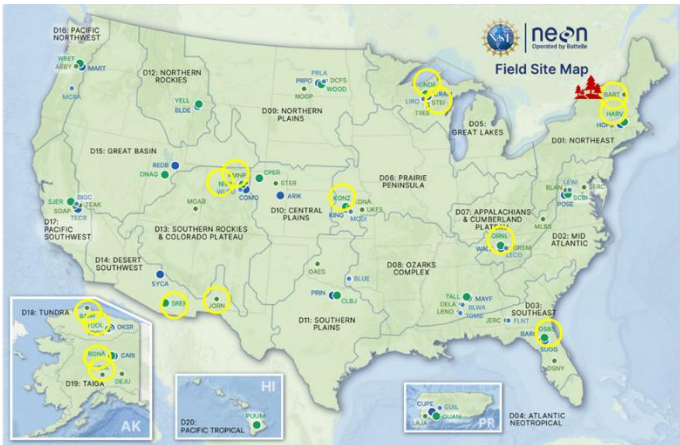
Andrew D. Richardson (ORCID: 0000-0002-0148-6714): Northern Arizona University, Center for Ecosystem Science and Society (ECOSS), AZ 86004, USA; Northern Arizona University, School of Informatics, Computing & Cyber Systems (SICCS), AZ 86004, USA.

Abstract

Long-term ecological data are essential for detecting impacts of climate change and other global change factors, and for making informed predictions about future change. However, long-term measurements are rarely replicated at the site level, which raises questions about their representativeness. We used a multiscale approach to evaluate the agreement of parallel observations from AmeriFlux and NEON (National Ecological Observatory Network) towers at Bartlett Experimental Forest, New Hampshire, USA. The two towers are separated by a horizontal distance of 93 m. We focused our analysis on standard meteorological

variables; fluxes of CO₂, sensible heat, and latent heat measured by eddy covariance; and phenology derived from PhenoCam imagery. Results suggest excellent agreement between AmeriFlux and NEON in meteorology and phenology, and good agreement in fluxes at the half-hourly scale. However, large disagreements in CO₂ and latent heat fluxes occurred at the annual scale, with implications especially for the forest carbon balance. The AmeriFlux tower measurements indicate a site that is close to carbon-neutral ($-8 \pm 65 \text{ g C m}^{-2} \text{ y}^{-1}$, mean $\pm 1 \text{ SD}$), whereas the NEON tower measurements indicate a forest that is a carbon sink ($-137 \pm 10 \text{ g C m}^{-2} \text{ y}^{-1}$). Causes of this disagreement may include measurement height (26 m vs. 35 m), which resulted in different flux footprints being measured by the two towers, and differences in the flux measurement systems. Our results suggest the need for caution when attempting to merge long-term flux data from two different measurement platforms, and when using measurements from any one measurement platform to inform decision-making on issues related to carbon accounting or natural climate solutions.

Graphical abstract



Sites where existing AmeriFlux towers are located near NEON sites

Each yellow circle represents a site pair, consisting of one NEON tower and its paired AmeriFlux tower. Circles are centered on the location of the NEON tower in each pair. Abbreviations: air temperature (Tair), soil temperature (Tsoil), photosynthetic photon flux density (PPFD), relative humidity (rH), actual vapor pressure (VP), vapor pressure deficit (VPD), net radiation (NETRAD), net shortwave radiation (NETSW), net longwave radiation (NETLW) and Enhanced Vegetation Index (EVI).

Continuity scorecard for Bartlett Experiment Forest

Excellent

Meteorology: Tair, Tsoil, NETRAD, NETSW
Phenology: GCC seasonality
Ecological similarity: canopy nitrogen

Good

Meteorology: PPFD, rH, VP, VPD
Phenology: transition dates
Ecological similarity: canopy height, EVI
Flux time domain: sensible heat flux
Flux frequency domain: CO₂ and sensible heat fluxes

Fair

Meteorology: NETLW
Flux time domain: CO₂ and latent heat fluxes
Flux frequency domain: latent heat fluxes



“Continuity scorecard” for the AmeriFlux and NEON datastreams at Bartlett Experimental Forest

Keywords

Carbon cycle, Eddy covariance, NEON Airborne Observation Platform (AOP), flux footprints, net ecosystem exchange, phenology, PhenoCam

1 Introduction

Long-term data are becoming increasingly valuable as climate change accelerates the occurrence of droughts, heat waves, and other climatological events that significantly impact ecosystem functioning (Lindenmayer et al., 2012). Long-term data are essential for detecting ecological changes and for understanding and predicting ecosystem response to climate change, as well as quantifying the impact of nature-based climate solutions (Delwiche et al., 2024). Eddy covariance flux measurements of CO₂, latent, and sensible heat, are particularly valuable for providing long-term insights into ecosystem function from hours to decades (Baldocchi, 2020; Delwiche et al., 2024). AmeriFlux has been collecting and coordinating eddy covariance data across the Americas since 1996 and provides a common platform for flux tower data sharing and collaboration (Novick et al., 2018). While some AmeriFlux data records are more than 25 y in length, most AmeriFlux sites operate on a 3-5 y time horizon (or less), contingent on grant funding, and none are guaranteed to continue for decades into the future. By comparison, NEON (the National Ecological Observatory Network) aims to provide high-quality ecological data on a continental scale using standardized protocols, over a 30-y period. However, NEON datasets were only initiated in 2017 (Nagy et al., 2021). The integration of new NEON data with pre-existing data from AmeriFlux represents a promising means of understanding global change impacts over longer time scales than either network might be capable of on its own. But it is not clear whether *existing AmeriFlux data can be used to extend, back in time, NEON measurements that did not begin until 2017*.

Integration of existing datasets from AmeriFlux with new data from NEON requires a high level of continuity between two data streams to minimize the likelihood of artifacts. For example, although AmeriFlux sites follow a common set of guidelines, the details of each eddy covariance flux measurement system, and the approaches to calculating fluxes from high-frequency measurements, differ across AmeriFlux sites, and between AmeriFlux and NEON sites. By comparison, all NEON sites use a common measurement protocol, an identical instrument configuration, and an identical data processing pipeline. Even if the flux measurement footprint were the same, the differences identified here could result in discrepancies in fluxes measured by two different measurement teams. Another important aspect is whether measurements conducted at different points are equally representative of the broader landscape; in heterogeneous ecosystems, this representativeness may largely determine the continuity between two data streams if the measurement footprints are different. Because flux measurements integrate across a footprint that extends for up to hundreds or thousands of meters in the upwind direction, it is generally assumed that the spatial domain being sampled is representative at the ecosystem scale (Burba, 2013). However, as it is uncommon for more than one eddy covariance measurement system to be established at a given site (exceptions include Hollinger et al. (2004); Poe et al. (2020); Post et al. (2015)), the assumption is rarely

tested and within-ecosystem heterogeneity can have a pronounced impact on measured fluxes (Chu et al., 2021; Oren et al., 2006).

Here, we conducted a multi-scale approach to evaluate the agreement between AmeriFlux and NEON data streams on both temporal and spatial scales at Bartlett Experimental Forest, New Hampshire, USA. The two towers are located 93 meters apart and share similar environmental conditions. This is the closest pairing among all sites with co-located NEON and AmeriFlux towers, making it a good starting point for comparing measurements from co-located towers. Specifically, we evaluated the agreement of:

- 1) meteorological data at the half-hourly scale: air and soil temperature (Tair and Tsoil), photosynthetic photon flux density (PPFD), net radiation (NETRAD), net shortwave radiation (NETSW), net longwave radiation (NETLW), relative humidity (rH), vapor pressure (VP) and vapor pressure deficit (VPD);
- 2) flux footprints, and ecological similarity of the flux footprints, based on leaf traits and canopy structure;
- 3) surface-atmosphere flux data at multiple time scales, namely at half-hourly, monthly, and annual scales.
- 4) surface-atmosphere flux data in frequency domain, i.e., wavelet power spectra of individual time series and wavelet coherence between paired flux data;
- 5) green chromatic coordinate (GCC) time series and transition dates derived from PhenoCam imagery.

Our results highlight the potential challenges of integration of new NEON data with pre-existing AmeriFlux data.

2 Methods

2.1 Study site

The Bartlett Experimental Forest (44°06'N, 71°3'W) is a temperate deciduous forest located in New Hampshire, within the White Mountain National Forest (Figure 1). The climate is humid continental with a mean annual temperature of approximately 6.6 °C. Temperatures range from -25 °C in winter to 30 °C in summer. Mean annual precipitation is 1300 mm, distributed evenly throughout the year. The mean stand age is roughly 100-125 years (Ouimette et al., 2018). Harvesting and silvicultural experiments have been conducted at Bartlett since the Experimental Forest was established in 1931, resulting in a heterogeneous patchwork of stand ages and species composition across the landscape. The dominant deciduous species are *Acer rubrum* (red maple; 29% basal area) and *Fagus grandifolia* (American beech, 25% basal area), with lesser amounts of other deciduous species such as *Betula alleghaniensis* (yellow birch, 9%), *Betula*

papyrifera (paper birch, 6%), and *Acer saccharum* (sugar maple, 5%). The most abundant evergreen conifer is *Tsuga canadensis* (Eastern hemlock, 14% basal area), while *Pinus strobus* (white pine) and *Picea rubens* (red spruce) are minor components of the forest. Canopy height is approximately 22 m within the tower footprint, according to a previous study (Ouimette et al., 2018) and NEON Airborne Observation Platform data.

The 26.5 m high AmeriFlux eddy covariance tower, operated by the USDA Forest Service and the University of New Hampshire, began flux measurements in January 2004 (Richardson and Hollinger, 2016), while the 35.7 m high NEON tower started flux measurements in January 2017. The two towers are located 93 m apart from each other (Figure 1c, d), in a broad u-shaped valley that slopes gently downhill (6% grade) to the north. In all directions, a ring of hills about 5 km distant rises approximately 700 m above the valley floor (Figure 1b).

The instruments installed on each tower are listed in Table S1. Turbulent fluxes of CO₂ (FCO₂) and sensible (H) and latent (LE) heat are measured at each tower using the eddy covariance method. However, measurement protocols regarding the sensors used, instrument type, measurement height, data acquisition frequency, flux processing, and quality control protocols all differ between the AmeriFlux and NEON tower (Table S1).

Our analysis of CO₂ exchange is based on FCO₂ rather than net ecosystem exchange, as the storage term is not measured at the AmeriFlux tower. While single-point storage correction could be made based on the CO₂ concentration measured at the height of flux measurements, this approach has been shown to poorly characterize diurnal variation in storage (Baldocchi, 2003). Meteorological and flux data from 2017 to 2022 for both sites were obtained from AmeriFlux BASE products (NEON, 2023a; Richardson and Hollinger, 2016).

2.2 Meteorological variables

We evaluated six major meteorological variables: air temperature (T_{air}), soil temperature (T_{soil}), photosynthetic photon flux density (PPFD), net radiation (NETRAD), net shortwave radiation (NETSW), net longwave radiation (NETLW), relative humidity (rH), vapor pressure deficit (VPD) and vapor pressure (VP). VPD and VP were calculated from half-hourly means of rH and T_{air}. T_{air}, T_{soil}, and PPFD are measured by sensors at multiple heights/depths on or adjacent to the NEON tower (Table S1), whereas the AmeriFlux tower has one primary sensor for each variable. The measurement heights and depths used in this analysis are reported in Table 1. For the meteorological variables (and later, for 30-minute fluxes), we evaluated agreement using (i) Type II regression using the R package lmodel2 (Pierre Legendre, 2008) to fit linear relationships between the paired NEON and AmeriFlux measurements; and (ii) using the mean, median, standard deviation, skewness and kurtosis of the difference between paired measurements.

2.3 Flux footprints

To estimate the surface area that contributes most to the measured turbulent flux, we calculated two-dimensional flux footprints using the Flux Footprint Prediction (FFP) model (Kljun et al., 2004), which is suitable for a wide range of boundary layer stratification conditions and measurement heights. Flux footprints vary by season and year, but in this study, we focused on peaking growing season, by computing the daytime and nighttime footprint climatology, respectively, for May, June, July in 2018. Roughness length and displacement height were estimated from mean canopy height (h) as $0.1h$, and $0.66h$, respectively. During our analysis, it became apparent that the sonic anemometer on the AmeriFlux tower had been installed upside-down for the duration of the period of comparison. We corrected the measured wind direction prior to footprint calculations. Wind roses showing the original wind directions and the corrected wind directions are provided in [Figure S1](#), suggesting better agreement with NEON wind directions after correction.

The flux footprint is described by a probability density function, providing information about the relative contribution (weight, %) of each point in the surrounding forest to the measured fluxes. Since the uncertainty of footprint models increase with distance upwind from receptor, we truncated the footprint climatology at the 80% cumulative contour of the source weights (Chu et al., 2021; Kljun et al., 2015). We focus on several leaf traits, e.g., canopy nitrogen (University Of New Hampshire Earth Systems Research Center, 2016) and height (NEON, 2023b), Enhanced Vegetation Index (EVI) (Moon et al., 2022) and vegetation area index (VAI) (Wang et al., 2024), to characterize “ecological similarity”. More details about the data products for canopy traits can be found in Table S2. We note that the values for this VAI product (mean value of 2.5) are lower than reported in other studies conducted at the same site. For example, Jenkins et al. (2007) reported a leaf area index (LAI) of 4.5 as measured by the optically based Li-Cor LAI-2000 instrument (average value across multiple plots), and LAI of 3.6 as measured by seasonal litterfall collection (Jenkins et al., 2007). We computed the footprint-weighted (i.e. locations within the footprint with a high source weight being weighted more heavily than locations with a low source weight, or locations outside of the 80% contour) distribution of the target traits of both towers. We used the Kolmogorov-Smirnov test (K-S test, Massey, 1951) to compare trait distributions, and a t-test to compare trait means.

2.4 Filtering, gap-filling, and flux partitioning

We filtered the flux data to remove outliers that were more than 6 times the interquartile range from the expected flux value (Liu et al., 2025; Richardson et al., 2019). To filter FCO_2 measurements during periods of insufficient turbulent mixing, we used the u^* threshold approach as implemented in *REddyProc* (Wutzler et al., 2018). Briefly, for FCO_2 , we estimated the u^* threshold distribution using a bootstrapping method (n

= 1000 samples), and then set our filtering threshold to the median (50th percentile, u^*_{50}) value. We tested different approaches to u^* filtering, first determining a single median u^* threshold value across all years (u^*_{single}), and then identifying the median u^* threshold separately for each year (u^*_{annual}). Finally, to systematically investigate the impact of the chosen u^* threshold on annual FCO₂ integrals, we filtered data using prescribed values for u^* of 0.2, 0.3, 0.4 and 0.5 m/s.

Input features for gapfilling flux data

To determine daily-to-annual flux integrals, we used the Extreme Gradient Boosting (XGBoost) machine learning algorithm for gap filling FCO₂, H and LE (Liu et al., 2025). As predictors, we used: Tair, Tsoil, VPD, PPFD, direct incoming PPFD (PPFD_DIR, $\mu\text{mol photon m}^{-2} \text{ s}^{-1}$), diffuse incoming PPFD (PPFD_DIF, $\mu\text{mol Photon m}^{-2} \text{ s}^{-1}$), and the Green Chromatic Coordinate (section 2.5) vegetation index from PhenoCam (Richardson et al., 2018). Missing predictors in the NEON dataset were filled (i) using measurements of the same variable measured at a different height on the NEON tower, or (ii) using the complementary measurement from the AmeriFlux tower, scaled using the linear models determined using the Model II regression already fit to the data (section 2.2). In order to validate the XGBoost gap filling for FCO₂, we also used an Artificial Neural Network (ANN) with the same input features. The procedures for running XGBoost and ANN for gap-filling can be found in Text S1. After gap-filling FCO₂ with XGBoost, we partitioned FCO₂ into ecosystem respiration (Reco) and gross canopy photosynthesis (GPP) using the nighttime approach within the *REddyProc* package (Reichstein et al., 2005; Wutzler et al., 2018).

2.5 Time domain comparison of 30-minute, monthly, and annual fluxes

Building on our analysis of meteorological data, we evaluated the agreement between simultaneous AmeriFlux and NEON 30-minute FCO₂, H, and LE measurements flux using Type II regression. We also aggregated gap-filled and partitioned fluxes to monthly and annual time steps for further comparison.

2.6 Frequency domain analysis of 30-minute fluxes

Wavelet spectral analysis

While 30-minute flux data are commonly analyzed in the time domain, they can also be analyzed in the frequency domain, which reveals how energy in the signal is distributed over a range of different temporal scales (Baldocchi et al., 2001). We used wavelet analysis (Katul et al., 2001; Stoy et al., 2005) to examine how (dis)agreement between 30-minute time series varies across temporal scales from sub-daily to multi-annual. There is a long history of using frequency-domain analyses to investigate phenomena occurring on very fast timescales, including turbulence (e.g., Farge, 1992; Kumar and Foufoula-Georgiou, 1997), but our analysis uses similar methods to address phenomena on much slower timescales, from 10⁰-10⁴ h.

The wavelet transform decomposes signals using dilated and translated functions called mother wavelets. In this study, the Morlet wavelet basis function was used as the mother wavelet, as it offers good time-frequency localization at high frequency resolution. Wavelet power was corrected to ensure a consistent definition of power across different time scales (Dietze et al., 2011; Liu et al., 2007). Gap-filled flux time series were normalized (mean=0 and standard deviation=1) before conducting wavelet analysis using the R package *WaveletComp* (Schmidbauer and Roesch, 2018). Wavelet spectra were evaluated down to the Nyquist frequency, equivalent to double the 30-minute sampling interval (i.e. 1 hour). Initially, we examined the wavelet spectra of individual time series. Due to the overall similarity observed in the spectral characteristics in paired observations, we then examined the wavelet spectra of the *difference* between two paired flux time series, where the *difference* was defined as $d = (\text{NEON} - \text{AmeriFlux})$.

Wavelet coherency

Wavelet analysis is particularly effective for non-stationary time series because a time series can be decomposed into time-frequency space (Torrence and Compo, 1998), providing information about how the different temporal scales of the signal change over time (Cazelles et al., 2008). A strength of this approach is that we can quantify significance in multiple timescales even when common power (i.e., shared variability) among time series on these scales is low (Stoy et al., 2013; Torrence and Compo, 1998). Wavelet coherence enabled determination of when, and on what timescales, AmeriFlux and NEON measurements are significantly related, and more importantly, when they are not.

2.7 Phenology

Both AmeriFlux and NEON towers at this site are equipped with phenology cameras facing north of the tower, to track seasonal changes in canopy color and structure (Richardson, 2019). The same model of camera is used on each tower, and cameras have been configured following the standard PhenoCam Network protocol (Richardson et al., 2018; Seyednasrollah et al., 2019). We used a standard vegetation index derived from camera data, the green chromatic coordinate (GCC_mean), calculated across a 3-day moving window, to track phenology. Full image processing details are described by Richardson *et al.*, 2018 (Richardson et al., 2018). We compared the GCC time series from the [bbc7](#) camera on the AmeriFlux tower with GCC time series from the [NEON.D01.BART.DP1.00033](#) camera on the NEON tower. Transition dates derived from these time series were determined as the first dates when the GCC time series reached 10% of its seasonal amplitude during GCC rising (spring) and falling (autumn) phases. Processed data were obtained from PhenoCam (<http://phenocam.nau.edu>).

3 Results

3.1 Meteorology

Meteorological variables demonstrated close agreement between AmeriFlux and NEON data streams for T_{air} and T_{soil} with R^2 larger than 0.99, but also with occasional notable outliers from the 1:1 line (Figure 2a,b). For PPFD, rH, and VPD, R^2 was greater than 0.97, but with more instances of outliers from the 1:1 line (Figure 2c-f). For PPFD, scattering occurred at low PPFD values ($< 500 \mu\text{mol m}^{-2} \text{s}^{-1}$), potentially due to PPFD sensors not being internally heated and thus sometimes covered by snow during winter; Shading, especially when solar elevation was low, may also be a factor. Obvious divergence from the 1:1 line was observed at high rH ($> 75\%$) and low VPD ($< 10 \text{ hPa}$).

Despite strong linear relationships for these variables, the confidence intervals of the regression slopes consistently did not include 1, and the confidence intervals of the regression intercepts did not include 0 (Table 1). The non-zero regression intercept indicates small but statistically significant biases for meteorological measurements between AmeriFlux and NEON, and the deviation of the regression slope from 1 indicates that this bias varied with the magnitude of the quantity being measured. Specifically, the bias between the AmeriFlux and NEON measurements varied between high and low values and hence over the course of the year. For example, at -10°C the AmeriFlux-measured T_{air} was generally cooler than NEON-measured T_{air} ($\approx -9.9^\circ\text{C}$) by 0.1°C , but at 30°C the AmeriFlux-measured T_{air} was warmer than the NEON-measured T_{air} ($\approx 29.8^\circ\text{C}$) by 0.2°C .

3.2 Flux footprints and leaf and canopy traits

Footprints of the AmeriFlux and NEON flux measurements generally overlapped but differed in size. The 80% NEON footprint climatology covered 0.56 km^2 , whereas the corresponding AmeriFlux footprint climatology, which was nested within the NEON footprint climatology, covered only 0.18 km^2 (Figure 3a, top panel). This difference in footprint size can be attributed to the higher flux measurement height for NEON (35.7 m) than AmeriFlux (26.5 m). Not surprisingly, the nighttime footprints were much larger than the corresponding daytime footprints, and the NEON nighttime footprint was 25 times larger than the AmeriFlux nighttime footprint (Figure 3a, bottom panel).

We used daytime footprint-weighted leaf and canopy traits to evaluate the ecological similarity of the flux footprints (Figure 3, Figure S2). We found significant differences in these traits (Figure 3), where the NEON footprint had lower canopy nitrogen (footprint weighted mean: 1.91% vs 1.93% for AmeriFlux), a taller canopy height (footprint-weighted mean: 17.5 m vs. 16.3 m for AmeriFlux) and higher VAI (footprint-weighted mean: $2.82 \text{ m}^2/\text{m}^2$ vs. $2.51 \text{ m}^2/\text{m}^2$ for AmeriFlux). Mean canopy EVI_{max} values were similar between the two footprints, although the frequency distributions differed significantly (K-S test, $P < 0.001$).

Seasonal patterns of albedo were similar for the two towers, but over the period June-August, the mean AmeriFlux albedo (0.167) was somewhat higher than the NEON albedo (0.141).

3.3 Fluxes in the time domain

30-minute fluxes

We found reasonably good agreement for FCO₂, and sensible (H) and latent (LE) heat flux, at the 30-minute time scale, but in all cases the confidence intervals on regression slopes did not include 1 and confidence intervals on intercepts did not include 0 (Table 1). The slope of FCO₂ between AmeriFlux and NEON data was very close to unity, 0.99, with R² value of 0.828 and root mean square error (RMSE) of 3.35 μmol m⁻² s⁻² (Figure 4), but the agreement of FCO₂ during the daytime (slope of 0.98, R² of 0.791, RMSE of 3.52 μmol m⁻² s⁻²) was much better than the agreement during the nighttime (slope of 0.97, R² of 0.497, RMSE of 2.36 μmol m⁻² s⁻², Figure S3). Of the three fluxes, the strongest correlation was observed for H (R² of 0.893, slope of 0.91, RMSE of 37.98 W m⁻²). By comparison, the correlation was lower, and the slope farther from 1, for LE (R² of 0.793, slope of 1.28, RMSE of 37.98 W m⁻²). Histograms of the difference in paired measurements (Figure 4d-4f, Table 1), showed relatively symmetric distributions for FCO₂ and H, but a skewed distribution for LE.

We found differences among FCO₂, H, and LE when 30-minute data were analyzed by year (Figure S4). The statistics of agreement for FCO₂ and H were similar across different years, but for LE the agreement was notably better in 2017 and 2018 (slope closest to 1:1) compared to other years (Figure S4c). We observed that the agreement for LE is significantly affected by rH (Figure S5). For rH > 60% (which occurred approximately two-thirds of time), the agreement becomes progressively worse as humidity increases (Figure S5b). The stickiness of water means that, compared to FCO₂, LE is more likely to be affected by multiple elements of system design (including tube length, heating, and flow rate), which could contribute to the observed differences, especially under higher-humidity conditions (Metzger et al., 2016; Zhang et al., 2023).

Monthly and annual flux integrals

Calculation of monthly or annual flux integrals requires gap-free data. Gap-filling statistics for XGBoost (Table S3) show generally similar levels of performance, in terms of RMSE, R², and bias for both NEON and AmeriFlux fluxes. However, one exception is nighttime FCO₂, and specifically the decrease in R² during nighttime compared to daytime, which is substantially more pronounced for the AmeriFlux (R² day = 0.873, R² night = 0.394 55% decrease) measurements than the NEON measurements (R² day = 0.852, R² night = 0.638; 25% decrease).

Monthly and annual FCO₂ integrals provide strongly contrasting views of the carbon (C) balance at Bartlett (Figure 5a). Whereas the AmeriFlux data indicate a forest that is, on average, close to C neutral (annual C balance ranging from -110 (sink) to +87 (source) g C m⁻² y⁻¹; mean ± 1 SD = 8 ± 65 g C m⁻² y⁻¹), the NEON data suggest a consistent C sink each year, with little interannual variability (annual C balance ranging from -152 (sink) to -125 (sink) g C m⁻² y⁻¹; mean ± 1 SD = -138 ± 10 g C m⁻² y⁻¹). These differences persist even when gap-filling is conducted using an alternative method based on ANN (Figure 5a; AmeriFlux mean ± 1 SD = -23 ± 30 g C m⁻² y⁻¹, NEON mean ± 1 SD = -179 ± 21 g C m⁻² y⁻¹). The stronger C sink measured by NEON from 2017 to 2022 is largely due to the accumulation of small FCO₂ differences throughout most of the year, and particularly during the winter, spring, and autumn months. From November through April, the NEON tower measured less net CO₂ release than the AmeriFlux tower every year. During the spring and autumn transition months of May and October, the NEON tower measured more net CO₂ uptake than the AmeriFlux tower every year. During the summer months of June through September, the patterns were not consistent; in some years, the NEON tower measured more net CO₂ uptake, and in other years the AmeriFlux tower did. The cumulative effect of these differences is that over the 6-year period, the NEON tower measured 780 g C m⁻² (130 g C m⁻² y⁻¹ on an annual basis) more uptake than the AmeriFlux tower (Figure S6).

For H, agreement in annual flux varied strongly by year; for the first three years, the annual H measured by NEON was greater than that measured by AmeriFlux, whereas for the last three years, the reverse was true (Figure 5b). In 2017 and 2018, much of the higher H measured by NEON can be attributed to substantially higher H measured during the non-growing season months (Figure 5b). In recent years, however, these differences largely disappeared, leading to improved agreement between H measured by AmeriFlux and NEON on both annual and monthly time scales (Figure 5b).

For LE, the annual flux was greater for the NEON measurements than the AmeriFlux measurements in each year, and especially from 2019-2022 when the difference exceeded 60% (Figure 5c). Differences in annual LE can be attributed to much higher monthly fluxes being measured by NEON during the peak growing season (Figure 5c).

Diel patterns and partitioned FCO₂

While the diel patterns for both AmeriFlux and NEON FCO₂ indicate strong photosynthetic uptake from June to September (Figure S7a), the two towers also show some uptake in May, and the NEON data show very modest uptake in April and October. Gap-filling with both XGB and ANN replicates these patterns (Figure 5, Figure S7). These differences observed in April and October may be the result of NEON's larger flux footprint, which includes a higher proportion of evergreen trees, enabling photosynthetic uptake during the shoulder seasons (see section 4.2c).

Pronounced differences in partitioned fluxes, GPP and Reco, were also observed between AmeriFlux and NEON. In two (2017, 2018) of 6 y study years, NEON annual GPP exceeded AmeriFlux annual GPP by over $50 \text{ g C m}^{-2} \text{ y}^{-1}$, while for other years the differences in annual GPP were $< 50 \text{ g C m}^{-2} \text{ y}^{-1}$ (Figure S8). There were no obvious patterns in the difference between monthly sums of GPP across different years. However, the Reco measured by NEON was higher than that measured by AmeriFlux for the first 3 y, with absolute differences of 3 to $128 \text{ g C m}^{-2} \text{ y}$. For the last 3 years, Reco measured by NEON was lower than that measured by AmeriFlux, with absolute differences ranging from 188 to $217 \text{ g C m}^{-2} \text{ y}^{-1}$.

3.4 Fluxes in the frequency domain

Wavelet spectra properties

The wavelet spectra for all three fluxes exhibit high spectral energy at the $\approx$ hourly time scales (periods when random errors and stochastic turbulent transport, including footprint variability, dominate), a spectral peak on the daily time scale (when diel cycles of temperature and solar radiation dominate), and another peak on the seasonal-to-annual scale (when periodicity of climate forcing is strong, Figure 6 abc). Notably, (a) for FCO_2 , there is more spectral energy in the NEON time series at the monthly scale and multiple-year scale (Figure 6a); (b) for H, the spectra for NEON and AmeriFlux are virtually identical across all timescales (Figure 6b); and (c) for LE, the spectral energy for the NEON time series is slightly higher than the AmeriFlux time series on the weekly-to-monthly scale and on the multiple-year scale (Figure 6c).

While these patterns might seem to contradict the finding that annual FCO_2 and LE sums were not in good agreement, this is consistent with our observation that differences in annual sums resulted from small differences (e.g. non-zero intercepts and slopes not equal to 1, in Table 1) accumulating over long periods. The spectral energy contributed by these small differences would be dwarfed by the much larger impacts of diel and annual forcing, which are largely driving the overall wavelet spectra. Indeed, there are prominent spectral peaks at the diel and monthly-to-annual time scales in the wavelet spectra of the *difference* between paired 30-minute flux measurements (Figure 6d), indicating persistent offsets between the NEON and AmeriFlux measurements on these timescales.

Wavelet coherence

Wavelet coherence analysis showed that there was strong coherence in H throughout the time series, whereas for FCO_2 and LE, the window of strong coherence (pink regions) is more limited to peak growing seasons (Figure 7). This may be explained by the fact that during peak growing seasons, the magnitude of fluxes is large, especially for FCO_2 and LE, and coherence becomes stronger. When the magnitude of fluxes is lower during winter, the coherence becomes weaker when the diel signal is less pronounced. In other words, on shorter time scales the observed patterns largely reflect seasonal changes in the underlying flux

signal and the noise (random measurement error) that is superimposed on that signal. When the signal is small (winter), and signal-to-noise ratio is low, then coherence is negligible.

For FCO₂, H, and LE, there was strong wavelet coherence at monthly and annual scales across the whole time series (Figure 7). At multi-year scales, the coherence occasionally becomes weaker. These observations are consistent with the change in agreement between the LE observations that were more closely aligned earlier in the study period, and the increasing trend in annual H for the AmeriFlux but not NEON measurements (Figure 5b).

3.5 Phenology

The time series of GCC from the AmeriFlux and the NEON phenocams were highly correlated, with R² of 0.977 (Figure 8a). Transition dates derived from six years of imagery were also in excellent agreement both in spring and autumn (Figure 8b). Differences in derived start- (SOS) and end-of-season (EOS) transition dates were small; NEON SOS was earlier than AmeriFlux SOS by 2.1 ± 0.4 days and NEON EOS was earlier than AmeriFlux EOS by 1.3 ± 1.2 days. Confidence intervals on the SOS and EOS transition dates, which are derived from the uncertainty in the smoothing spline used for transition date extraction, range (across years and cameras) range from ± 2 –4 days for SOS, and ± 2 –5 for EOS. Thus, the differences in transition dates are within the ranges of uncertainty for a single measurement. The difference in growing season length (days between SOS and EOS) was negligible, -1.0 ± 2.7 days.

4 Discussion

We quantified timescales and conditions under which the meteorological, flux, and phenology measurements from the AmeriFlux and NEON towers at Bartlett Experimental Forest are in (dis)agreement. These towers are separated by a distance of less than 100 m and therefore experience essentially identical weather. They experience similar turbulence environments when the wind is along the axis between the two towers. But, the higher measurement height on the NEON tower results in a three-fold larger footprint than is measured by the AmeriFlux tower. Key canopy traits (both footprint-weighted mean values and distributions) such as canopy nitrogen, canopy height, and VAI varied significantly between the two towers, although the difference in footprint-weighted canopy nitrogen (0.02%) may be too small to be ecologically relevant. Our two-tower comparison is notably different from that presented Hollinger et al. (2004), where the separation between towers was larger and the footprints were generally non-overlapping.

We have summarized our results into a report card, as shown in the graphical abstract. Here, we will elaborate on the paired measurements that show good agreement and highlight the discrepancies. We also consider the question, “*Can NEON and AmeriFlux data be merged to provide backward (before NEON measurements were initiated) or forward (if AmeriFlux measurements were to be discontinued) continuity?*”

4.1 Measurements in good agreement

Based on the analysis of half-hourly data, meteorological variables measured at multiple heights for NEON, such as Tair, Tsoil, and PPFD, showed the best agreement when the measurement heights or depths were similar ($R^2 > 0.99$, slope close to 1). For other meteorological variables without a vertical profile, such as rH, VP, and VPD, we compared NEON measurements at approximately 35 m with AmeriFlux measurements taken at around 22-26 m. Despite the 10 m difference in measurement height, the agreement remained very good ($R^2 > 0.97$). Other metrics of agreement, including mean, median, and standard deviation of the difference between paired measurements (Table 1) confirm this assessment, and suggest that the measurement differences are only slightly higher than typical instrument specifications (e.g. $\pm 0.2^\circ\text{C}$ and 2% rH for Vaisala HMP-45 humidity sensor; $\pm 5\%$ for LI-COR LI-190 quantum sensor).

The agreement between paired 30-minute flux measurements is not as strong, with lower correlation between NEON and AmeriFlux measurements (Figure 4, Table 1). We found the correlation was highest for H ($R^2 = 0.89$), and substantially lower for both FCO₂ ($R^2 = 0.83$) and LE ($R^2 = 0.79$). Consistent with our results, Hollinger et al. (2004) also found that, for 30-min fluxes measured at a pair of towers in homogeneous forest separated by ≈ 775 m, the correlation was strongest for H ($R^2 = 0.90$), followed by FCO₂ ($R^2 = 0.79$) and LE ($R^2 = 0.75$). One reason for this pattern may be that the high-frequency random errors in H are generally smaller than those for LE or FCO₂ (Hollinger and Richardson, 2005). And, as H is measured using data from just the sonic anemometer, differences in the design and configuration of the infrared gas analysis measurement systems do not affect measurements of H, where as they do impact measurements of FCO₂ and LE (Schmidt et al., 2012).

This assessment of agreement should be considered in light of the potential to measure different quantities accurately and precisely. For example, in addition to ± 0.2 - 0.4°C accuracy of air temperature sensors, errors associated with aspiration of the sensor are an additional source of uncertainty. Flux measurements are subject to substantial random (predominantly due to the stochastic nature of turbulence) and systematic (associated with high- and low-frequency losses, limitations of the measurement system design, and imperfect corrections) errors (Richardson et al., 2012). For FCO₂, random errors of ≈ 10 - 20% in 30-minute data can accumulate to an uncertainty in annual sums of $\pm 50 \text{ g C m}^{-2} \text{ y}^{-1}$ (Richardson et al., 2012; Schmidt et al., 2012). In our analysis, the standard deviation of the difference between paired flux measurements exceeds by about 50% of what would be expected based on random measurement alone (Hollinger and Richardson, 2005; Richardson et al., 2006), while the mean and median difference between paired flux measurements indicate systematic biases (Figure 4) in the 30-minute measurements that could further accumulate to substantial differences on longer timescales. These systematic biases most likely result from the aggregate effects of many small differences in how the high-frequency data are measured, and how

fluxes are calculated, corrected, and quality-controlled. While nothing can be done at this point to rectify differences in the underlying measurement systems, applying identical processing to the high-frequency data could *potentially* result in a reduction in systematic biases. However, we note that it is extremely rare for end-users of publicly-available flux data to reprocess the raw data themselves, and hence we have focused our evaluation of agreement on the data as currently available through the AmeriFlux BASE data portal.

4.2 Measurements with disagreement

4.2.1 Carbon dioxide fluxes at monthly to annual time scales

The AmeriFlux measurements indicate a forest that is nearly C neutral, whereas the NEON measurements indicate a C sink. These differences have substantial implications for forest C cycling, carbon accounting, and nature-based climate solutions. We explore several hypotheses to explain the observed differences.

a) Gap filling. The fact that agreement in 30-minute measurements appeared to be generally good ($R^2 = 0.838$; [Figure 4a](#)), with what (visually) appears to be minimal bias (regression line almost indistinguishable from the 1:1 line), but annual FCO_2 sums were vastly different, suggests that gap-filling may have played a potential role. Using ANN instead of XGBoost to fill gaps resulted in slight changes (but not improved agreement) in the patterns of interannual variability ([Figure 5a](#)), although annual FCO_2 still differed by $156 \pm 38 \text{ g C m}^{-2} \text{ y}^{-1}$ (vs. $130 \pm 69 \text{ g C m}^{-2} \text{ y}^{-1}$ using XGB), allowing us to rule out the choice of gap filling algorithm as an explanation. However, the distribution of gaps in the FCO_2 record varied depending on time of day and time of year, and between the two towers. Briefly (for a more detailed analysis, see [Text S2](#)), we found that mean fluxes were most similar (mean difference of approximately $0.25 \mu\text{mol m}^{-2} \text{ s}^{-1}$, equivalent to about 5 g C m^{-2} per 1000 measurements) when there were valid measurements from both towers (case 1, [Table S4](#)) or when both towers had a missing measurement that required gap filling (case 4, [Table S4](#)). Larger discrepancies (mean difference of approximately $0.50 \mu\text{mol m}^{-2} \text{ s}^{-1}$, equivalent to about 10 g C m^{-2} per 1000 measurements) occurred when one tower had valid measurement and the other did not (cases 2 and 3, [Table S4](#)). Differences were more than twice as large during the day (mean difference, across all cases, of $0.49 \mu\text{mol m}^{-2} \text{ s}^{-1}$) than at night (mean difference, across all cases, of $0.19 \mu\text{mol m}^{-2} \text{ s}^{-1}$). Thus, the fluxes measured by the two towers tend to be offset by a small but significant amount, as indicated by either $\Delta F = -0.28 \mu\text{mol m}^{-2} \text{ s}^{-1}$ for case 1 in [Table S4](#), or the y-axis intercept of $-0.33 \mu\text{mol m}^{-2} \text{ s}^{-1}$ in [Figure 4a](#). Uneven training data (differences in the distribution of gaps, under different cases by season and day/night) further contribute to the differences in annual FCO_2 as the inherent differences in measurements are propagated through to differences in gap-filled values. We conclude that ultimately, small but persistent differences in half-hourly fluxes are accumulating to become substantial differences in annual FCO_2 : the overall $\Delta F = -0.34 \mu\text{mol m}^{-2} \text{ s}^{-1}$ (across all cases in [Table S4](#)) equates to the observed difference in annual

C balance of $130 \text{ g C m}^{-2} \text{ y}^{-1}$ even though a bias of this magnitude is imperceptible in Figure 4a. Recent advances in flux mapping (Rey-Sanchez et al., 2022; Xu et al., 2017) could potentially help improve agreement between the two towers by providing a more direct comparison of fluxes for their overlapping footprint areas. Moreover, for most years, the number of valid nighttime measurements during the summer months (May–September) is often below 100, whereas winter months generally have at least twice as many nighttime observations, based on a u^* threshold of 0.3 (Figure S9). This means that there are more nighttime observations that require filling during the summer, but also that goodness-of-fit statistics for the XGB gap-filling may be skewed towards nighttime performance during winter, when fluxes are negligible (Figure S9). This low summer nighttime data coverage poses challenges for gap-filling during peak growing seasons.

b) Topography and drainage. The total annual ecosystem respiration measured by NEON was substantially lower than that measured by AmeriFlux (Figure S6a; the cumulative difference is about 370 g C m^{-2} , or $62 \text{ g C m}^{-2} \text{ y}^{-1}$ on an annual basis). The difference in annual ecosystem respiration can be attributed to the small difference in nighttime fluxes, $0.19 \mu\text{mol m}^{-2} \text{ s}^{-1}$, which accumulated to a total nighttime ΔFCO_2 (and hence ecosystem respiration) difference of $36 \text{ g C m}^{-2} \text{ y}^{-1}$ (Table S4; these $36 \text{ g C m}^{-2} \text{ y}^{-1}$ explain about 28% of the overall difference in estimated sink strength between the two towers). The remaining $26 \text{ g C m}^{-2} \text{ y}^{-1}$ difference in ecosystem respiration (i.e., total difference of $62 \text{ g C m}^{-2} \text{ y}^{-1}$ less the nighttime difference of $36 \text{ g C m}^{-2} \text{ y}^{-1}$) must therefore arise from the extrapolation of the nighttime measurements to estimate daytime ecosystem respiration, as we implemented the nighttime partitioning method.. Then, because the difference between Reco and FCO_2 was used to estimate GPP, this means that the difference between GPP between the NEON and AmeriFlux tower (ΔGPP) was driven by both ΔReco and ΔFCO_2 , with ΔFCO_2 dominating because of its much large variance, as shown by the strong correlation between 30-minute ΔGPP with ΔFCO_2 ($R^2 = 0.832$, Figure S6c)

The landscape at Bartlett Experimental Forest slopes gradually downward within and surrounding the flux footprint, potentially leading to cold air drainage and underestimation of nighttime respiration (Goulden et al., 1996). We used a relatively high u^* threshold to exclude stable conditions and minimize the impact of drainage. We do not have data that would enable quantification of differences in nocturnal drainage (or horizontal advection more generally) between the two towers, but it may be that the greater measurement height (and hence potentially larger storage flux), and/or generally larger footprint, of the NEON tower results in greater susceptibility to under-estimation of nighttime fluxes, even with u^* filtering.

b) u^* threshold selection. In the results reported above, we used a single u^* threshold across all years for each tower. Surprisingly, the u^*_{50} threshold determined using REddyProc was very similar for AmeriFlux (0.286 m/s) and NEON (0.294 m/s). Previous studies have provided mixed guidance on whether u^* should

be fixed or whether it should vary over time (Barr et al., 2013; Ouimette et al., 2018). We found that the agreement in annual FCO₂ was no better when u^* was estimated separately for each year (Figure S9e). We also evaluated whether different fixed u^* thresholds could improve the agreement in annual NEE (Figure S9c). A higher u^* threshold consistently resulted in a weaker estimated C sink, but regardless of the value of u^* chosen, the patterns of interannual variability in annual FCO₂ were in poor agreement between the two towers.

c) Flux footprints and differences in biological activity. The overlap in flux footprints between NEON and AmeriFlux was only about 30%. Across the climatological flux footprint of each tower, mean footprint-weighted canopy nitrogen was slightly lower for the NEON tower than the AmeriFlux tower, whereas VAI was higher for the NEON tower than the AmeriFlux tower (Figure 3). The distribution of footprint-weighted EVI max differed between the two towers, as the NEON footprint includes more regions with high EVI_{max} (and also high VAI, Figure 3d,e) that are generally outside the AmeriFlux tower footprint and contribute to differences in the distribution of each trait. The difference in VAI does not appear to be associated with increased GPP within the NEON footprint, but we note that increased midday C uptake during April and October (Figure S6) for the NEON tower suggest a greater influence of evergreen conifers relative to the footprint of the AmeriFlux tower. Indeed, the footprint-weighted percentage of mixed forest components was higher for NEON (13.3%) than for AmeriFlux (2.0%; Figure S10). Thus, differences in biological activity within the flux footprint of each tower could explain some of the difference in annual C sink strength in this mixed temperate forest. But, although management activities conducted to the west of both towers in 2018 and 2021 were expected to have had different influences on each tower's FCO₂ measurements, we found no evidence for this (Figure S11). To contextualize these footprint differences, we note that in a previous study, Oren et al. (2006) found that even within a relatively uniform, monospecific pine plantation, spatial variability in ecosystem activity (including plants and decomposers) and leaf area index contributed to half of the uncertainty in annual CO₂ flux measurements (Oren et al., 2006). And, consistent with our comparison of two towers with nearly 10 m in height difference, Schmid et al. (2003) found that the annual net CO₂ uptake of a deciduous forest varied by 33% ± 17% depending on the measurement height on the same tower (34 m vs. 46 m) and the associated footprint differences (Schmid et al., 2003). However, in that study, NEE of CO₂ was more negative (stronger sink) for the measurements at 34 m compared to 46 m—i.e., the opposite of what we observed between the AmeriFlux and NEON towers at Bartlett.

4.2.2 Latent heat fluxes at annual time scales

We observed progressively worse agreement in annual LE over the 6 years of this analysis (Figure 5c). While our focus here has been to evaluate the agreement or continuity between NEON and AmeriFlux data

streams, rather than to judge the quality of individual data streams, this pattern would appear to point to previously unrecognized data quality issues for LE measured on the AmeriFlux tower, particularly from 2019 onward. Indeed, energy balance closure became progressively worse for the AmeriFlux measurements over this time (Figure S12e). Similar findings of declining energy balance closure over time have been reported by Stoy et al. (2013) and are attributed to aging inlet tubes that attenuate high-frequency fluctuations, leading to underestimation of LE (Stoy et al., 2013). The AmeriFlux site PIs (coauthors DYH, APO, and ADR) thus recommend against using the 2019-2022 LE data until a more thorough investigation has been conducted.

More generally, these results highlight the importance of regular and rigorous evaluation of data quality, and careful attention to maintenance to ensure sustained, long-term consistency. For decades, AmeriFlux Tech Team's site visits, which involve installation of a "roving" system (PECS, or Portable Eddy Covariance System) adjacent to existing tower measurements, has been the means by which data quality and consistency is ensured across the network. The Tech Team's site visit to the Bartlett AmeriFlux tower in 2016 found good agreement against the roving system for LE ($R^2 = 0.94$, slope = 1.03, intercept = -3.8 W m^{-2}). However, it was also noted at that time (Chan, unpublished report) "Spectral analysis illustrated significant signal attenuation of the gas scalars, particularly water vapor, due to the length and possibly age of the inlet tube used on the *in situ* eddy covariance system." While the final fluxes (including spectral corrections) were found to generally match those from the PECS, the relatively large (20-30%) spectral corrections required for LE are an additional source of uncertainty. On top of this, maintenance trips to the AmeriFlux tower were minimal in 2019 and challenged by COVID and personnel changes in later years. Deferral of what had previously been regular maintenance activities (e.g. annual replacement of the 25 m inlet tube) may reasonably be assumed to be playing a role in dramatic under-measurement of LE at monthly to annual timescales. (Note that although it is standard for the LI-7200 to mount at the top of the tower, with a very short inlet tube, for consistency with LI-6262 measurements conducted over the first 10 y at Bartlett, the LI7200 is housed in the shed at the base of the tower, necessitating a long inlet tube).

4.2.3 Missing data as a quality metric

One objective criterion by which data quality may be judged is the fraction of missing data. For meteorological variables (Figure S13b), except for a high fraction of gaps in NEON vapor pressure data, the fraction of missing values for other meteorological variables for most of the years was <10%. For fluxes, data gaps are much more common (see section 4.2). Overall, 40-60% of 30-minute flux measurements were missing, but this varied by flux, time of year, and across years (Figure S13, Figure S14). NEON flux data had large gaps earlier in the study period (2017-2019), but these have largely disappeared in recent years;

in both 2018 and 2021, three-quarters of the year's FCO₂ measurements were missing, and this was mostly because of long gaps due to power and pump issues. Long gaps introduce more errors in gap-filling (Liu, 2024), leading to larger uncertainties in annual sums. The continuity of flux data measurements at both towers could certainly be improved, and future efforts to minimize gaps in the flux data would undoubtedly contribute to reduced uncertainties.

4.3 Potential for integrating AmeriFlux and NEON datastreams at Bartlett Experimental Forest

With AmeriFlux and NEON operating towers in close proximity, and above similar vegetation, at more than a half-dozen sites, some site PIs (including some coauthors of this paper) have wondered whether their own measurements are still needed or whether they are now redundant. The overarching motivation for this study was to determine whether overlapping AmeriFlux and NEON time series were sufficiently consistent to be stitched, spliced, or aligned to yield a single unified FCO₂ time series longer than either of the two original time series. From our analysis, it would be misguided to simply append the NEON FCO₂ time series to the AmeriFlux FCO₂ time series. Re-scaling the data from one site to match the other site is possible, but these scaling relationships appear to vary from year-to-year (Figure S4), and there is little or no synchronicity in the interannual variability of annual FCO₂, GPP, or Reco (Figures 5a and Figure S4). Finally, the 130 g C m⁻² y⁻¹ difference between the near-neutral C balance indicated by the AmeriFlux tower measurements, and the CO₂ sink indicated by the NEON tower measurements, is approximately equal in magnitude to the measured long-term wood growth of 170-200 g C m⁻² y⁻¹ (Ouimette et al., 2018; Teets et al., 2022). Such differences are large enough it may not be possible for them to be reconciled without considerable additional effort; one potential starting point would be to reprocess the high-frequency data from both towers using a standardized data processing pipeline.

Notably, one of the key goals of NEON is to provide standardized measurements, using protocols and measurement systems that are as similar as possible across all sites in the network (Nagy et al., 2021), from low-stature tundra at high elevations and high latitudes, to old-growth conifer forests in the Pacific Northwest and tropical ecosystems in Puerto Rico and Hawaii. NEON's towers were never intended to provide long-term continuity for existing AmeriFlux measurements. Better agreement between the AmeriFlux and NEON FCO₂ measurements at Bartlett likely could have been achieved only by compromising NEON's efforts at standardization.

Importantly, there does appear to be much greater continuity in the meteorological measurements made on the two towers. It may be reasonable to use the first 13 years (2004-2016, predating NEON) of meteorological and environmental data from the AmeriFlux tower to supplement the 2017-present NEON measurements. One potential application would provide locally measured driver data for a forest ecosystem

model or other process-based model. This approach could also potentially drive, in hindcasting mode, an empirical or machine learning model trained on the 2017-2022 NEON flux data.

5 Conclusion

We observed strong agreement of meteorological and phenological measurements from the AmeriFlux and NEON towers at Bartlett Experimental Forest. Good agreement was also noted in sensible heat flux at both half-hourly and annual scales. However, despite apparently good agreement in latent heat flux and CO₂ flux at the half-hourly scale, large discrepancies emerged at the annual scale. Annual latent heat flux differed by 460 MJ m⁻² y⁻¹, while annual FCO₂ differed by about 130 g C m⁻² y⁻¹. These results, made from two towers separated by only 93 m, highlight the need for careful evaluation when using tower flux measurements for carbon accounting, assessment of natural climate solutions, and climate-smart forest management practices. Future work will expand this analysis framework to additional sites across nine NEON domains where NEON and AmeriFlux towers can be paired, with separation distances from <1 km to 75 km. This will help further evaluate the consistency and representativeness of NEON and AmeriFlux data across different ecosystems, and across sites using a range of measurement systems and data processing protocols.

Data availability statement

- The measurement heights for different sensors are downloaded from <https://ameriflux.lbl.gov/data/measurement-height/>.
- PhenoCams data (green chromatic coordinate and transition dates) are downloaded from <https://phenocam.nau.edu/webcam/>.
- Eddy covariance data are downloaded from AmeriFlux BASE data products: <https://ameriflux.lbl.gov/data/download-data/>.
- Canopy height map is downloaded from <https://data.neonscience.org/data-products/DP3.30015.001>.
- Canopy Nitrogen Map of year 2012 is downloaded from <https://knb.ecoinformatics.org/view/doi%3A10.5063%2FF1V9860T>.
- EVI_{max} map is based on high spatial resolution land surface phenology dataset https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=2033.

- Vegetation area index (VAI) from the NEON Discrete-Return LiDAR Point Cloud are downloaded from [Data Portal - Data Package Summary | Environmental Data Initiative \(EDI\) \(edirepository.org\)](https://data.neon.org/DataPortal/DataPackageSummary/EnvironmentalDataInitiative(EDI)/edirepository.org)

Data and code availability

https://github.com/YujieLiu666/NEON_AmeriFlux_continuity

Data and code will be uploaded to Zenodo with a DOI upon acceptance of the manuscript.

Acknowledgement

The work is supported by the National Science Foundation under awards 2105828, 2224545, and 2222437. PhenoCam data used in this research were provided by the PhenoCam Network, which has been supported by the National Science Foundation, the Long-Term Agroecosystem Research (LTAR) network which is supported by the United States Department of Agriculture (USDA), the U.S. Department of Energy, the U.S. Geological Survey, the Northeastern States Research Cooperative, and the USA National Phenology Network. Field measurements at the Bartlett AmeriFlux site have been supported by the USDA Forest Service's Northern Research Station, the National Science Foundation through support for the Hubbard Brook Long Term Ecological Research (LTER) program (awards 1114804 and 1637685), and the Northeastern States Research Cooperative (awards 12-DG-11242307-065 and 23-DG-11242311-020). This material is based in part upon work supported by the National Ecological Observatory Network (NEON), a program sponsored by the U.S. National Science Foundation (NSF) and operated under cooperative agreement by Battelle. YL also thanks Emma Gibbons Reich and Kimberly Samuels-Crow for the discussions about basic concepts for wavelet analysis. ChatGPT was used by YL to check grammar and spelling but was not used to generate any manuscript text.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

References

- Baldocchi, D., Falge, E., Wilson, K., 2001. A spectral analysis of biosphere–atmosphere trace gas flux densities and meteorological variables across hour to multi-year time scales. *Agricultural and Forest Meteorology* 107, 1–27. [https://doi.org/10.1016/S0168-1923\(00\)00228-8](https://doi.org/10.1016/S0168-1923(00)00228-8)
- Baldocchi, D., Novick, K., Keenan, T., Torn, M., 2024. AmeriFlux: Its Impact on our understanding of the ‘breathing of the biosphere’, after 25 years. *Agricultural and Forest Meteorology* 348, 109929. <https://doi.org/10.1016/j.agrformet.2024.109929>
- Baldocchi, D.D., 2020. How eddy covariance flux measurements have contributed to our understanding of *Global Change Biology*. *Global Change Biology* 26, 242–260. <https://doi.org/10.1111/gcb.14807>
- Baldocchi, D.D., 2003. Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future. *Global Change Biology* 9, 479–492. <https://doi.org/10.1046/j.1365-2486.2003.00629.x>
- Barr, A.G., Richardson, A.D., Hollinger, D.Y., Papale, D., Arain, M.A., Black, T.A., Bohrer, G., Dragoni, D., Fischer, M.L., Gu, L., Law, B.E., Margolis, H.A., McCaughey, J.H., Munger, J.W., Oechel, W., Schaeffer, K., 2013. Use of change-point detection for friction–velocity threshold evaluation in eddy-covariance studies. *Agricultural and Forest Meteorology* 171–172, 31–45. <https://doi.org/10.1016/j.agrformet.2012.11.023>
- Burba, G., 2013. Eddy Covariance Method for Scientific, Industrial, Agricultural and Regulatory Applications: A Field Book on Measuring Ecosystem Gas Exchange and Areal Emission Rates. LI-COR Biosciences. <https://doi.org/10.13140/RG.2.1.4247.8561>
- Cazelles, B., Chavez, M., Berteaux, D., Ménard, F., Vik, J.O., Jenouvrier, S., Stenseth, N.C., 2008. Wavelet analysis of ecological time series. *Oecologia* 156, 287–304. <https://doi.org/10.1007/s00442-008-0993-2>
- Chu, H., Luo, X., Ouyang, Z., Chan, W.S., Dengel, S., Biraud, S.C., Torn, M.S., Metzger, S., Kumar, J., Arain, M.A., Arkebauer, T.J., Baldocchi, D., Bernacchi, C., Billesbach, D., Black, T.A., Blanken, P.D., Bohrer, G., Bracho, R., Brown, S., Brunsell, N.A., Chen, J., Chen, X., Clark, K., Desai, A.R., Duman, T., Durden, D., Fares, S., Forbrich, I., Gamon, J.A., Gough, C.M., Griffis, T., Helbig, M., Hollinger, D., Humphreys, E., Ikawa, H., Iwata, H., Ju, Y., Knowles, J.F., Knox, S.H., Kobayashi, H., Kolb, T., Law, B., Lee, X., Litvak, M., Liu, H., Munger, J.W., Noormets, A., Novick, K., Oberbauer, S.F., Oechel, W., Oikawa, P., Papuga, S.A., Pendall, E., Prajapati, P., Prueger, J., Quinton, W.L., Richardson, A.D., Russell, E.S., Scott, R.L., Starr, G., Staebler, R., Stoy, P.C., Stuart-Haëntjens, E., Sonnentag, O., Sullivan, R.C., Suyker, A., Ueyama, M., Vargas, R., Wood, J.D., Zona, D., 2021. Representativeness of Eddy-Covariance flux footprints for areas surrounding AmeriFlux sites. *Agricultural and Forest Meteorology* 301–302, 108350. <https://doi.org/10.1016/j.agrformet.2021.108350>
- Delwiche, K.B., Nelson, J., Kowalska, N., Moore, C.E., Shirkey, G., Tarin, T., Cleverly, J.R., Keenan, T.F., 2024. Charting the Future of the FLUXNET Network. *Bulletin of the American Meteorological Society* 105, E466–E473. <https://doi.org/10.1175/BAMS-D-23-0316.1>
- Dietze, M.C., Vargas, R., Richardson, A.D., Stoy, P.C., Barr, A.G., Anderson, R.S., Arain, M.A., Baker, I.T., Black, T.A., Chen, J.M., Ciais, P., Flanagan, L.B., Gough, C.M., Grant, R.F., Hollinger, D., Izaurralde, R.C., Kucharik, C.J., Lafleur, P., Liu, S., Lokupitiya, E., Luo, Y., Munger, J.W., Peng, C., Poulter, B., Price, D.T., Ricciuto, D.M., Riley, W.J., Sahoo, A.K., Schaefer, K., Suyker, A.E., Tian, H., Tonitto, C., Verbeeck, H., Verma, S.B., Wang, W., Weng, E., 2011. Characterizing the performance of ecosystem models across time scales: A spectral analysis of the North American Carbon Program site-level synthesis. *J. Geophys. Res.* 116, G04029. <https://doi.org/10.1029/2011JG001661>
- Farge, M., 1992. Wavelet transforms and their applications to turbulence. *Annu. Rev. Fluid Mech.* 24, 395–458. <https://doi.org/10.1146/annurev.fl.24.010192.002143>

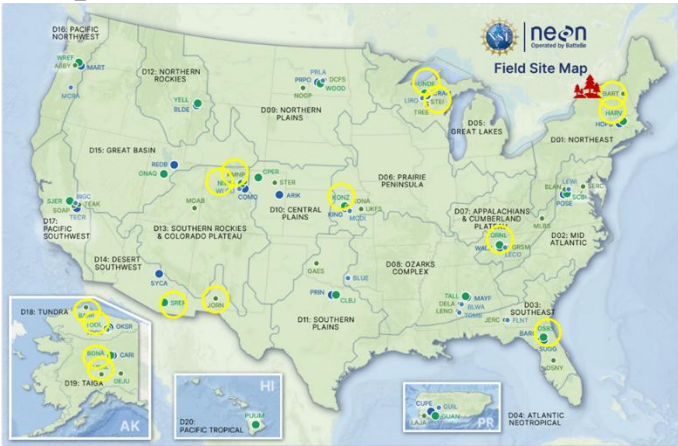
- Goulden, M.L., Munger, J.W., Fan, S., Daube, B.C., Wofsy, S.C., 1996. Measurements of carbon sequestration by long-term eddy covariance: methods and a critical evaluation of accuracy. *Global Change Biology* 2, 169–182. <https://doi.org/10.1111/j.1365-2486.1996.tb00070.x>
- Hollinger, D.Y., Aber, J., Dail, B., Davidson, E.A., Goltz, S.M., Hughes, H., Leclerc, M.Y., Lee, J.T., Richardson, A.D., Rodrigues, C., Scott, N.A., Achuatavari, D., Walsh, J., 2004. Spatial and temporal variability in forest-atmosphere CO₂ exchange. *Global Change Biology* 10, 1689–1706. <https://doi.org/10.1111/j.1365-2486.2004.00847.x>
- Hollinger, D.Y., Richardson, A.D., 2005. Uncertainty in eddy covariance measurements and its application to physiological models. *Tree Physiology* 25, 873–885. <https://doi.org/10.1093/treephys/25.7.873>
- Jenkins, J.P., Richardson, A.D., Braswell, B.H., Ollinger, S.V., Hollinger, D.Y., Smith, M.-L., 2007. Refining light-use efficiency calculations for a deciduous forest canopy using simultaneous tower-based carbon flux and radiometric measurements. *Agricultural and Forest Meteorology* 143, 64–79. <https://doi.org/10.1016/j.agrformet.2006.11.008>
- Katul, G., Lai, C.-T., Schäfer, K., Vidakovic, B., Albertson, J., Ellsworth, D., Oren, R., 2001. Multiscale analysis of vegetation surface fluxes: from seconds to years. *Advances in Water Resources* 24, 1119–1132. [https://doi.org/10.1016/S0309-1708\(01\)00029-X](https://doi.org/10.1016/S0309-1708(01)00029-X)
- Kljun, N., Calanca, P., Rotach, M.W., Schmid, H.P., 2015. A simple two-dimensional parameterisation for Flux Footprint Prediction (FFP). *Geosci. Model Dev.* 8, 3695–3713. <https://doi.org/10.5194/gmd-8-3695-2015>
- Kljun, N., Calanca, P., Rotach, M.W., Schmid, H.P., 2004. A Simple Parameterisation for Flux Footprint Predictions. *Boundary-Layer Meteorology* 112, 503–523. <https://doi.org/10.1023/B:BOUN.0000030653.71031.96>
- Kumar, P., Foufoula-Georgiou, E., 1997. Wavelet analysis for geophysical applications. *Reviews of Geophysics* 35, 385–412. <https://doi.org/10.1029/97RG00427>
- Lindenmayer, D.B., Likens, G.E., Andersen, A., Bowman, D., Bull, C.M., Burns, E., Dickman, C.R., Hoffmann, A.A., Keith, D.A., Liddell, M.J., Lowe, A.J., Metcalfe, D.J., Phinn, S.R., Russell-Smith, J., Thurgate, N., Wardle, G.M., 2012. Value of long-term ecological studies. *Austral Ecology* 37, 745–757. <https://doi.org/10.1111/j.1442-9993.2011.02351.x>
- Liu, Y., 2024. Robust filling of extra-long gaps in eddy covariance CO₂ flux measurements from a temperate deciduous forest using eXtreme Gradient Boosting. <https://doi.org/10.5281/ZENODO.14027040>
- Liu, Y., Lucas, B., Bergl, D.D., Richardson, A.D., 2025. Robust filling of extra-long gaps in eddy covariance CO₂ flux measurements from a temperate deciduous forest using eXtreme Gradient Boosting. *Agricultural and Forest Meteorology* 364, 110438. <https://doi.org/10.1016/j.agrformet.2025.110438>
- Liu, Y., San Liang, X., Weisberg, R.H., 2007. Rectification of the bias in the wavelet power spectrum. *Journal of Atmospheric and Oceanic Technology* 24, 2093–2102. <https://doi.org/10.1175/2007JTECHO511.1>
- Metzger, S., Burba, G., Burns, S.P., Blanken, P.D., Li, J., Luo, H., Zulueta, R.C., 2016. Optimization of an Enclosed Gas Analyzer Sampling System for Measuring Eddy Covariance Fluxes of H₂O and CO₂. *Atmos. Meas. Tech.* 9, 1341–1359. <https://doi.org/10.5194/amt-9-1341-2016>
- Moon, M., Richardson, A.D., Milliman, T., Friedl, M.A., 2022. A high spatial resolution land surface phenology dataset for AmeriFlux and NEON sites. *Sci Data* 9, 448. <https://doi.org/10.1038/s41597-022-01570-5>
- Nagy, R.C., Balch, J.K., Bissell, E.K., Cattau, M.E., Glenn, N.F., Halpern, B.S., Ilankoon, N., Johnson, B., Joseph, M.B., Marconi, S., O’Riordan, C., Sanovia, J., Swetnam, T.L., Travis, W.R., Wasser, L.A., Woolner, E., Zarnetske, P., Abdulrahim, M., Adler, J., Barnes, G., Bartowitz, K.J., Blake, R.E., Bombaci, S.P., Brun, J., Buchanan, J.D., Chadwick, K.D., Chapman, M.S., Chong, S.S., Chung, Y.A., Corman, J.R., Couret, J., Crispo, E., Doak, T.G., Donnelly, A., Duffy, K.A., Dunning, K.H., Duran, S.M., Edmonds, J.W., Fairbanks, D.E., Felton, A.J., Florian, C.R., Gann, D., Gebhardt, M., Gill, N.S., Gram, W.K., Guo, J.S., Harvey, B.J., Hayes, K.R., Helmus, M.R.,

- Hensley, R.T., Hondula, K.L., Huang, T., Hundertmark, W.J., Iglesias, V., Jacinthe, P., Jansen, L.S., Jarzyna, M.A., Johnson, T.M., Jones, K.D., Jones, M.A., Just, M.G., Kaddoura, Y.O., Kagawa-Vivani, A.K., Kaushik, A., Keller, A.B., King, K.B.S., Kitzes, J., Koontz, M.J., Kouba, P.V., Kwan, W., LaMontagne, J.M., LaRue, E.A., Li, D., Li, B., Lin, Y., Liptzin, D., Long, W.A., Mahood, A.L., Malloy, S.S., Malone, S.L., McGlinchy, J.M., Meier, C.L., Melbourne, B.A., Mietkiewicz, N., Morisette, J.T., Moustapha, M., Muscarella, C., Musinsky, J., Muthukrishnan, R., Naithani, K., Neely, M., Norman, K., Parker, S.M., Perez Rocha, M., Petri, L., Ramey, C.A., Record, S., Rossi, M.W., SanClements, M., Scholl, V.M., Schweiger, A.K., Seyednasrollah, B., Sihi, D., Smith, K.R., Sokol, E.R., Spaulding, S.A., Spiers, A.I., St. Denis, L.A., Staccone, A.P., Stack Whitney, K., Stanitski, D.M., Stricker, E., Surasinghe, T.D., Thomsen, S.K., Vasek, P.M., Xiaolu, L., Yang, D., Yu, R., Yule, K.M., Zhu, K., 2021. Harnessing the NEON data revolution to advance open environmental science with a diverse and data-capable community. *Ecosphere* 12, e03833. <https://doi.org/10.1002/ecs2.3833>
- NEON, (National Ecological Observatory Network), 2023a. AmeriFlux US-xBR NEON Bartlett Experimental Forest (BART), URL: <https://www.osti.gov/servlets/purl/1579542>. <https://doi.org/10.17190/AMF/1579542>
- NEON, (National Ecological Observatory Network), 2023b. Discrete Return LiDAR Point Cloud (DP1.30003.001), URL: <https://data.neonscience.org/data-products/DP1.30003.001>.
- Novick, K.A., Biederman, J.A., Desai, A.R., Litvak, M.E., Moore, D.J.P., Scott, R.L., Torn, M.S., 2018. The AmeriFlux network: A coalition of the willing. *Agricultural and Forest Meteorology* 249, 444–456. <https://doi.org/10.1016/j.agrformet.2017.10.009>
- Oren, R., Hsieh, C., Stoy, P., Albertson, J., McCarthy, H.R., Harrell, P., Katul, G.G., 2006. Estimating the uncertainty in annual net ecosystem carbon exchange: spatial variation in turbulent fluxes and sampling errors in eddy-covariance measurements. *Global Change Biology* 12, 883–896. <https://doi.org/10.1111/j.1365-2486.2006.01131.x>
- Ouimette, A.P., Ollinger, S.V., Richardson, A.D., Hollinger, D.Y., Keenan, T.F., Lepine, L.C., Vadeboncoeur, M.A., 2018. Carbon fluxes and interannual drivers in a temperate forest ecosystem assessed through comparison of top-down and bottom-up approaches. *Agricultural and Forest Meteorology* 256–257, 420–430. <https://doi.org/10.1016/j.agrformet.2018.03.017>
- Pierre Legendre, 2008. lmodel2: Model II Regression. <https://doi.org/10.32614/CRAN.package.lmodel2>
- Poe, J., Reed, D.E., Abraha, M., Chen, J., Dahlin, K.M., Desai, A.R., 2020. Geospatial coherence of surface-atmosphere fluxes in the upper Great Lakes region. *Agricultural and Forest Meteorology* 295, 108188. <https://doi.org/10.1016/j.agrformet.2020.108188>
- Post, H., Hendricks Franssen, H.J., Graf, A., Schmidt, M., Vereecken, H., 2015. Uncertainty analysis of eddy covariance CO₂ flux measurements for different EC tower distances using an extended two-tower approach. *Biogeosciences* 12, 1205–1221. <https://doi.org/10.5194/bg-12-1205-2015>
- Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., Valentini, R., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biology* 11, 1424–1439. <https://doi.org/10.1111/j.1365-2486.2005.001002.x>
- Rey-Sanchez, C., Arias-Ortiz, A., Kasak, K., Chu, H., Szutu, D., Verfaillie, J., Baldocchi, D., 2022. Detecting Hot Spots of Methane Flux Using Footprint-Weighted Flux Maps. *JGR Biogeosciences* 127, e2022JG006977. <https://doi.org/10.1029/2022JG006977>
- Richardson, A.D., 2019. Tracking seasonal rhythms of plants in diverse ecosystems with digital camera imagery. *New Phytologist* 222, 1742–1750. <https://doi.org/10.1111/nph.15591>

- Richardson, A.D., Aubinet, M., Barr, A.G., Hollinger, D.Y., Ibrom, A., Lasslop, G., Reichstein, M., 2012. Uncertainty Quantification, in: Aubinet, M., Vesala, T., Papale, D. (Eds.), *Eddy Covariance*. Springer Netherlands, Dordrecht, pp. 173–209. https://doi.org/10.1007/978-94-007-2351-1_7
- Richardson, A.D., Hollinger, D., 2016. AmeriFlux US-Bar Bartlett Experimental Forest. <https://doi.org/10.17190/AMF/1246030>
- Richardson, A.D., Hollinger, D.Y., Burba, G.G., Davis, K.J., Flanagan, L.B., Katul, G.G., William Munger, J., Ricciuto, D.M., Stoy, P.C., Suyker, A.E., Verma, S.B., Wofsy, S.C., 2006. A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes. *Agricultural and Forest Meteorology* 136, 1–18. <https://doi.org/10.1016/j.agrformet.2006.01.007>
- Richardson, A.D., Hollinger, D.Y., Shoemaker, J.K., Hughes, H., Savage, K., Davidson, E.A., 2019. Six years of ecosystem-atmosphere greenhouse gas fluxes measured in a sub-boreal forest. *Sci Data* 6, 117. <https://doi.org/10.1038/s41597-019-0119-1>
- Richardson, A.D., Hufkens, K., Milliman, T., Aubrecht, D.M., Chen, M., Gray, J.M., Johnston, M.R., Keenan, T.F., Klosterman, S.T., Kosmala, M., Melaas, E.K., Friedl, M.A., Frolking, S., 2018. Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery. *Sci Data* 5, 180028. <https://doi.org/10.1038/sdata.2018.28>
- Schmid, H.P., Su, H.-B., Vogel, C.S., Curtis, P.S., 2003. Ecosystem-atmosphere exchange of carbon dioxide over a mixed hardwood forest in northern lower Michigan. *J. Geophys. Res.* 108, 2002JD003011. <https://doi.org/10.1029/2002JD003011>
- Schmidbauer, H., Roesch, A., 2018. WaveletComp 1.1: A guided tour through the R package. R package version 1.1.
- Schmidt, A., Hanson, C., Chan, W.S., Law, B.E., 2012. Empirical assessment of uncertainties of meteorological parameters and turbulent fluxes in the AmeriFlux network. *J. Geophys. Res.* 117, 2012JG002100. <https://doi.org/10.1029/2012JG002100>
- Syednasrollah, B., Adam M. Young, Young, A.M., Hufkens, K., Milliman, T., Friedl, M.A., Frolking, S., Richardson, A.D., 2019. Tracking vegetation phenology across diverse biomes using Version 2.0 of the PhenoCam Dataset. *Scientific Data* 6, 222–222. <https://doi.org/10.1038/s41597-019-0229-9>
- Stoy, P.C., Dietze, M.C., Richardson, A.D., Vargas, R., Barr, A.G., Anderson, R.S., Arain, M.A., Baker, I.T., Black, T.A., Chen, J.M., Cook, R.B., Gough, C.M., Grant, R.F., Hollinger, D.Y., Izaurrealde, R.C., Kucharik, C.J., Lafleur, P., Law, B.E., Liu, S., Lokupitiya, E., Luo, Y., Munger, J.W., Peng, C., Poulter, B., Price, D.T., Ricciuto, D.M., Riley, W.J., Sahoo, A.K., Schaefer, K., Schwalm, C.R., Tian, H., Verbeeck, H., Weng, E., 2013. Evaluating the agreement between measurements and models of net ecosystem exchange at different times and timescales using wavelet coherence: an example using data from the North American Carbon Program Site-Level Interim Synthesis. *Biogeosciences* 10, 6893–6909. <https://doi.org/10.5194/bg-10-6893-2013>
- Stoy, P.C., Katul, G.G., Siqueira, M.B.S., Juang, J.-Y., McCarthy, H.R., Kim, H.-S., Oishi, A.C., Oren, R., 2005. Variability in net ecosystem exchange from hourly to inter-annual time scales at adjacent pine and hardwood forests: a wavelet analysis. *Tree Physiology* 25, 887–902. <https://doi.org/10.1093/treephys/25.7.887>
- Teets, A., Moore, D.J.P., Alexander, M.R., Blanken, P.D., Bohrer, G., Burns, S.P., Carbone, M.S., Ducey, M.J., Fraver, S., Gough, C.M., Hollinger, D.Y., Koch, G., Kolb, T., Munger, J.W., Novick, K.A., Ollinger, S.V., Ouimette, A.P., Pederson, N., Ricciuto, D.M., Syednasrollah, B., Vogel, C.S., Richardson, A.D., 2022. Coupling of tree growth and photosynthetic carbon uptake across six north American forests. *JGR Biogeosciences* 127. <https://doi.org/10.1029/2021JG006690>
- Torrence, C., Compo, G.P., 1998. A Practical Guide to Wavelet Analysis. *Bull. Amer. Meteor. Soc.* 79, 61–78. [https://doi.org/10.1175/1520-0477\(1998\)079%253C0061:APGTWA%253E2.0.CO;2](https://doi.org/10.1175/1520-0477(1998)079%253C0061:APGTWA%253E2.0.CO;2)
- University Of New Hampshire Earth Systems Research Center, 2016. Canopy Nitrogen, Bartlett Experimental Forest, NH USA, 2012. <https://doi.org/10.5063/F1V9860T>
- Wang, J., Choi, D.H., LaRue, E., Atkins, J.W., Foster, J.R., Matthes, J.H., Fahey, R.T., Fei, S., Hardiman, B.S., 2024. 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>

- Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Šigut, L., Menzer, O., Reichstein, M., 2018. Basic and extensible post-processing of eddy covariance flux data with REddyProc (preprint). *Biogeochemistry: Air - Land Exchange*. <https://doi.org/10.5194/bg-2018-56>
- Xu, K., Metzger, S., Desai, A.R., 2017. Upscaling tower-observed turbulent exchange at fine spatio-temporal resolution using environmental response functions. *Agricultural and Forest Meteorology* 232, 10–22. <https://doi.org/10.1016/j.agrformet.2016.07.019>
- Zhang, W., Jung, M., Migliavacca, M., Poyatos, R., Miralles, D.G., El-Madany, T.S., Galvagno, M., Carrara, A., Arriga, N., Ibrom, A., Mammarella, I., Papale, D., Cleverly, J.R., Liddell, M., Wohlfahrt, G., Markwitz, C., Mauder, M., Paul-Limoges, E., Schmidt, M., Wolf, S., Brümmer, C., Arain, M.A., Fares, S., Kato, T., Ardö, J., Oechel, W., Hanson, C., Korkiakoski, M., Biraud, S., Steinbrecher, R., Billesbach, D., Montagnani, L., Woodgate, W., Shao, C., Carvalhais, N., Reichstein, M., Nelson, J.A., 2023. The effect of relative humidity on eddy covariance latent heat flux measurements and its implication for partitioning into transpiration and evaporation. *Agricultural and Forest Meteorology* 330, 109305. <https://doi.org/10.1016/j.agrformet.2022.109305>

Graphical abstract



Sites where existing AmeriFlux towers are located near NEON sites

Each yellow circle represents a site pair, consisting of one NEON tower and its paired AmeriFlux tower. Circles are centered on the location of the NEON tower in each pair. Abbreviations: air temperature (Tair), soil temperature (Tsoil), photosynthetic photon flux density (PPFD), relative humidity (rH), actual vapor pressure (VP), vapor pressure deficit (VPD), net radiation (NETRAD), net shortwave radiation (NETSW), net longwave radiation (NETLW) and Enhanced Vegetation Index (EVI).

Continuity scorecard for Bartlett Experiment Forest	
Excellent	
Meteorology: Tair, Tsoil, NETRAD, NETSW Phenology: GCC seasonality Ecological similarity: canopy nitrogen	
Good	
Meteorology: PPFD, rH, VP, VPD Phenology: transition dates Ecological similarity: canopy height, EVI Flux time domain: sensible heat flux Flux frequency domain: CO ₂ and sensible heat fluxes	
Fair	
Meteorology: NETLW Flux time domain: CO ₂ and latent heat fluxes Flux frequency domain: latent heat fluxes	



“Continuity scorecard” for the AmeriFlux and NEON datastreams at Bartlett Experimental Forest

Tables

Table 1: Half-hourly meteorological, flux data, and their agreement for AmeriFlux and NEON measurements (2017-2022) at Bartlett Forest, New Hampshire, USA. Several of these variables are measured at multiple heights (Tair, PPFD) or depths (Tsoil) at the NEON tower. For these three variables, we report statistics for measurements from the individual sensor that was in closest agreement (highest R^2 , slope closest to 1, and intercept closest to 0) with corresponding measurement on the AmeriFlux tower. The 95% confidence intervals (2.5th and 97.5th percentiles) for the intercept and slope estimates are calculated using Model II regression for paired measurements between AmeriFlux and NEON. Slopes, intercepts and R^2 values are given in [Figure 2](#) and [Figure 4](#).

	Tair	Tsoil	PPFD	rH	VP	VPD	NETRAD	NETSW	NETLW	FCO2	H	LE
Unit	°C	°C	$\mu\text{mol m}^{-2} \text{s}^{-1}$	%	kPa	kPa	W m^{-2}	W m^{-2}	W m^{-2}	$\mu\text{mol m}^{-2} \text{s}^{-1}$	W m^{-21}	W m^{-21}
<u>Measurement height (m)</u>												
AmeriFlux	23.80	-0.10	23.80	22.30	26.50	22.30	25.00	25.00	25.00	23.80	23.80	23.80
NEON	25.50	-0.07	26.11	35.19	35.19	35.19	35.67	35.67	35.67	35.19	35.19	35.19
<u>Evaluation metrics</u>												
RMSE	0.32	0.60	45.78	5.60	0.11	1.02	19.54	17.21	15.69	3.35	37.98	47.41
R^2	0.999	0.991	0.992	0.971	0.988	0.976	0.991	0.995	0.864	0.828	0.893	0.793
2.5%-Intercept	0.035	0.051	4.223	-2.435	0.008	0.461	-3.648	0.785	0.160	-0.365	0.264	9.942
97.5%-Intercept	0.038	0.061	4.550	-2.289	0.009	0.470	-3.548	0.919	0.437	-0.297	0.548	10.689
2.5%-Slope	0.992	0.968	1.026	0.972	0.921	1.033	1.032	1.027	1.128	0.983	0.910	1.277
97.5%-Slope	0.993	0.970	1.027	0.974	0.922	1.036	1.033	1.028	1.134	0.997	0.916	1.293
<u>Differences (NEON - AmeriFlux)</u>												
Mean	-0.02	-0.21	12.15	-4.27	-0.06	0.06	-0.93	4.44	-5.56	-0.28	-3.60	23.48
Median	0.01	0.18	0.49	-3.61	-0.03	0.03	-2.70	1.08	-5.86	-0.30	-1.46	11.89
Mode	0.01	-0.18	0.12	-2.44	-1.77	0.00	5.07	-0.68	-11.86	-23.56	-342.78	-196.11
Standard deviation	0.32	0.56	44.13	3.63	0.08	0.07	19.51	16.63	14.67	3.34	37.81	41.19
Skewness	3.90	-0.10	3.70	-1.15	-1.97	2.13	1.29	1.84	0.84	-0.17	-0.41	1.40
Kurtosis	294.36	0.44	73.50	4.03	14.92	10.94	17.83	26.13	26.45	3.25	8.50	4.03

Abbreviations: air temperature (Tair), soil temperature (Tsoil), photosynthetic photon flux density (PPFD), relative humidity (rH), vapor pressure (VP), vapor pressure deficit (VPD), net radiation (NETRAD), net shortwave radiation (NETSW) and net longwave radiation (NETLW).

Figures

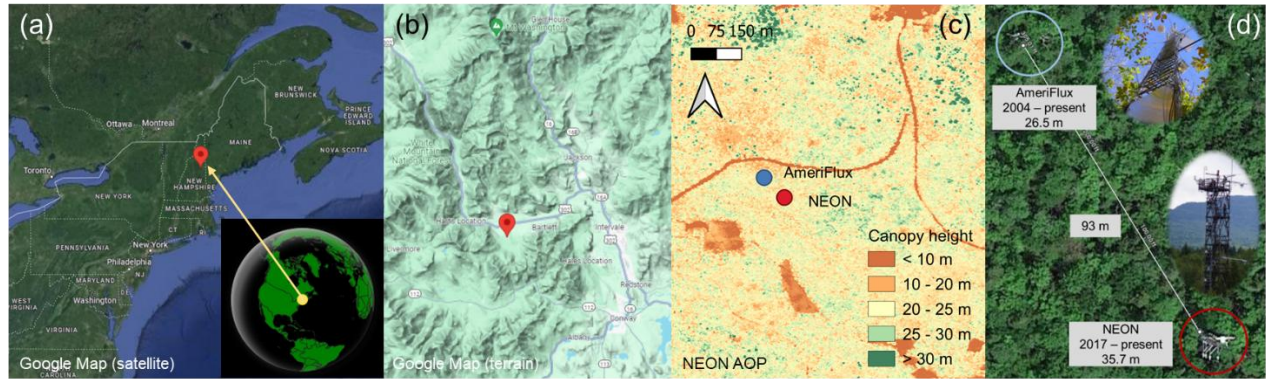


Figure 1: Location of Bartlett Experiment Forest, (a) Google Map (satellite), (b) Google Map (terrain), (c) NEON AOP, canopy height model from LiDAR point cloud, (d) locations of AmeriFlux and NEON eddy covariance towers, showing flux measurement heights.

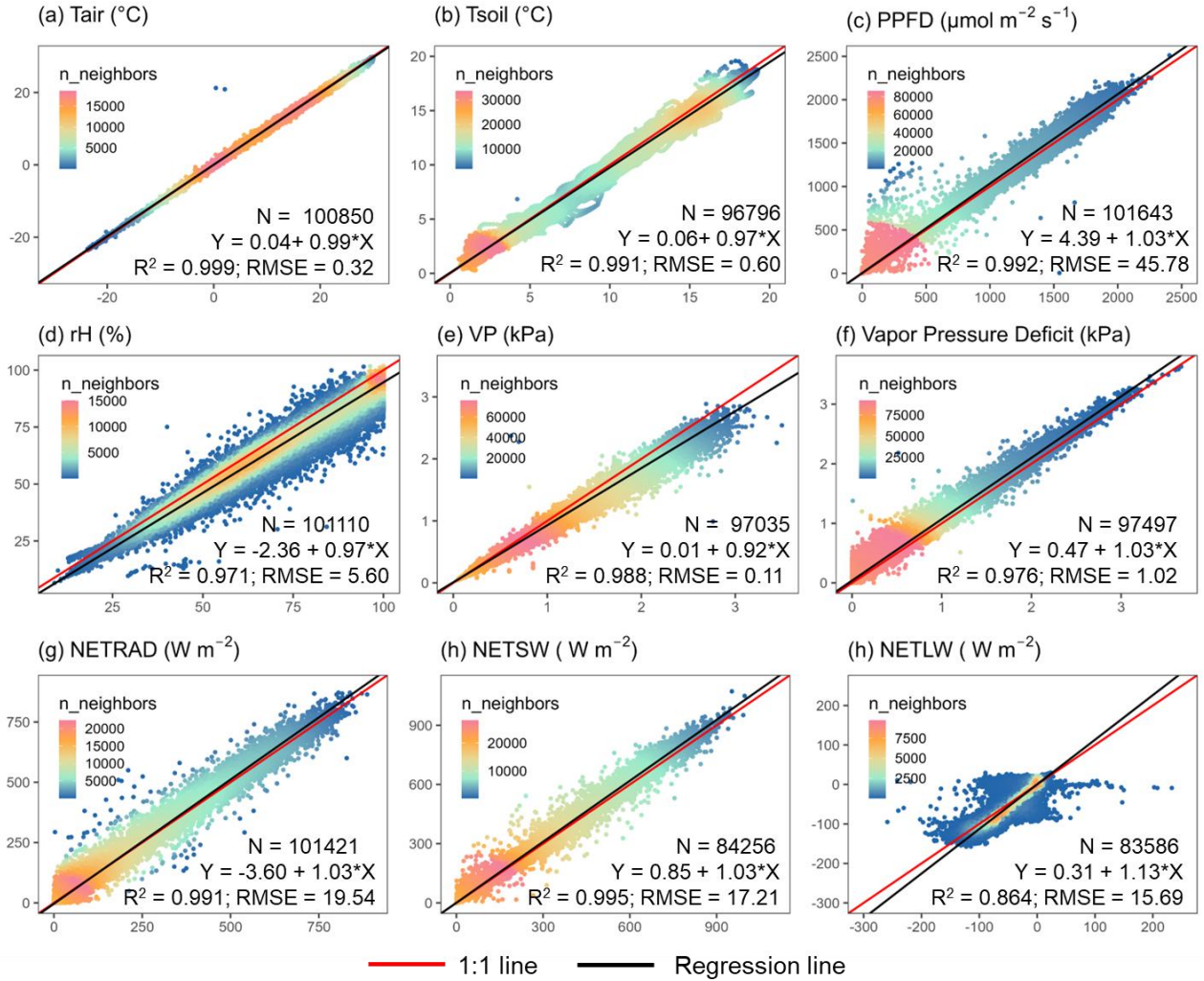


Figure 2: Scatter density plots of simultaneous measurements of half-hourly data for: (a) air temperature (Tair), (b) soil temperature (Tsoil), (c) photosynthetic photon flux density (PPFD), (d) relative humidity (rH), (e) vapor pressure (VP), (f) vapor pressure deficit (VPD), (g) net radiation (NETRAD), (h) net shortwave radiation (NETSW) and (i) net longwave radiation (NETLW) measured by AmeriFlux and NEON towers at Bartlett Experimental Forest. Individual points are colored by the number of neighboring points ($n_neighbors$). The R^2 , slope and intercept between variables are evaluated using Model II regression. The measurement heights and confidence intervals for the intercept and slope can be found in Table 1. A two-channel rather than four-channel, net radiometer is used at the AmeriFlux site, and thus while NETLW is measured it is not made available in the AmeriFlux BASE publication. Thus, for this analysis, NETLW was computed as NETRAD - NETSW.

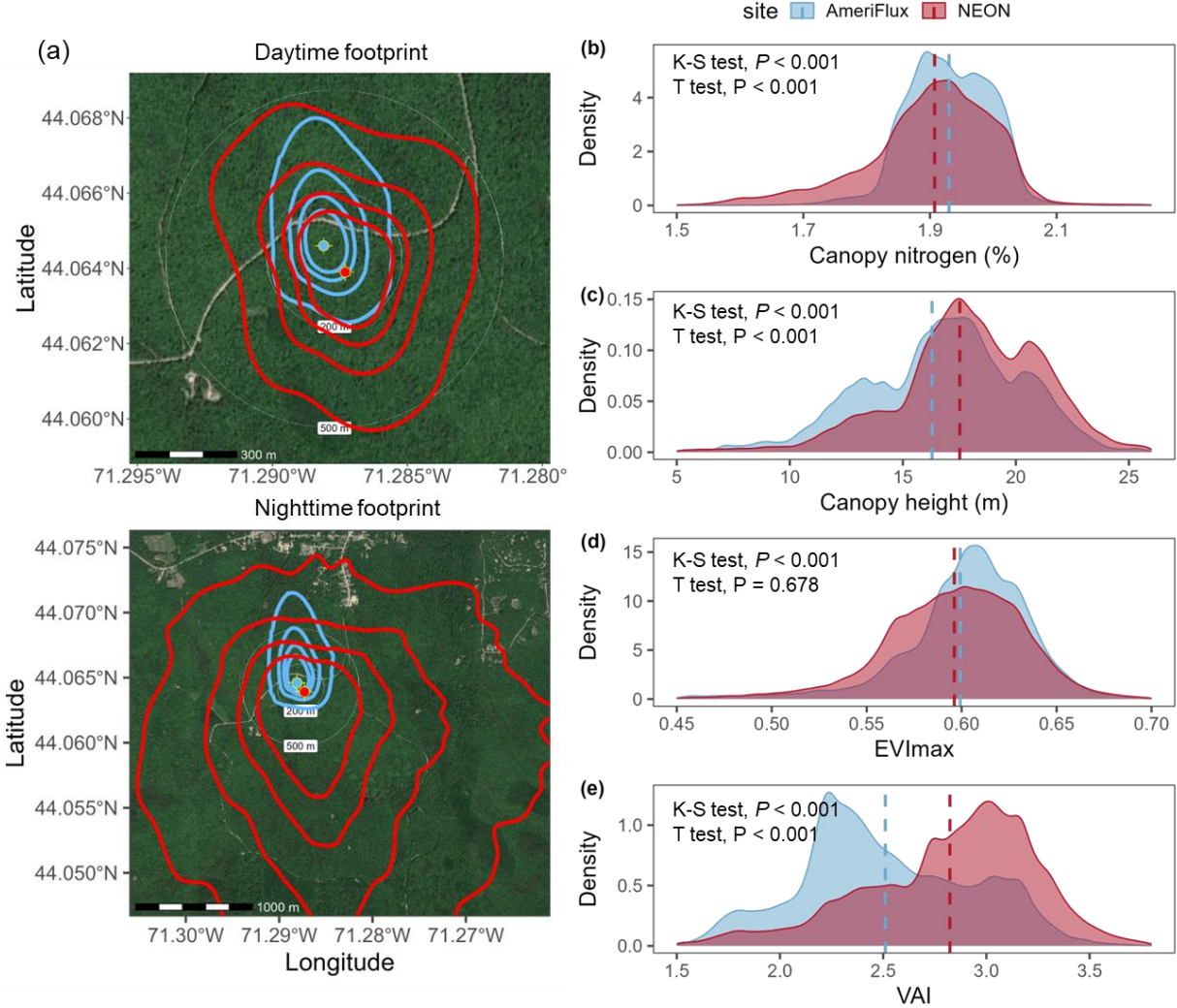


Figure 3: (a) Footprint climatology (50-80% cumulative contour lines) of AmeriFlux and NEON towers overlaid on a true-color google satellite image for daytime footprint (top panel) and nighttime footprint (bottom panel) respectively (Note that the map scale differs between the daytime and nighttime footprints); and ecological similarities (i.e., leaf and canopy traits within each flux footprint with 80% cumulative contour) evaluated using (b) canopy nitrogen (%), (d) canopy height (m) and (e) maximum Enhanced Vegetation Index (EVI_{max}) and (e) Vegetation Area Index (VAI). The Kolmogorov-Smirnov test (K-S test) was used to evaluate whether the two samples come from the same underlying distribution or if they differ significantly in their distributions.

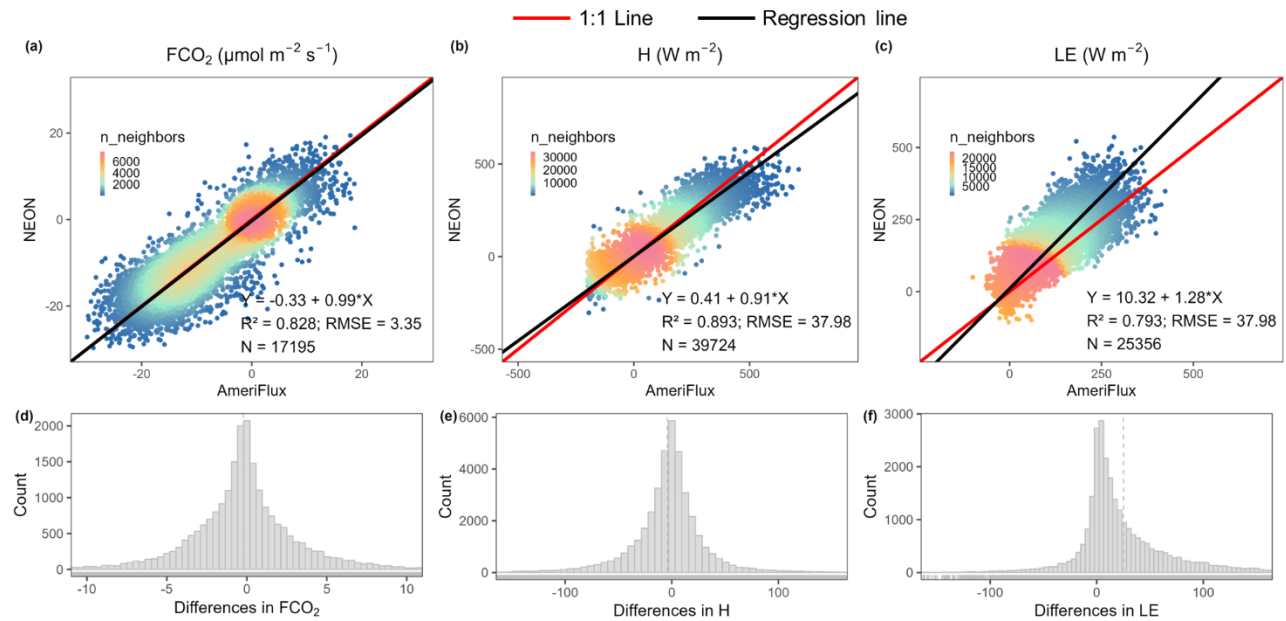


Figure 4: Top panels: scatter density plot comparing simultaneous measurements of half-hourly data: (a) CO₂ flux (FCO₂), (b) sensible heat flux (H), and (c) latent heat flux (LE) from the AmeriFlux and NEON measurements at Bartlett Experimental Forest. Bottom panels: histograms of differences in fluxes (NEON – AmeriFlux) for (d) FCO₂, (e) H and (f) LE respectively. The R^2 , slope and intercept between variables are evaluated using Model II regression. Individual points are colored by the number of neighboring points ($n_neighbors$). Data for each year are plotted in [Figure S2](#).

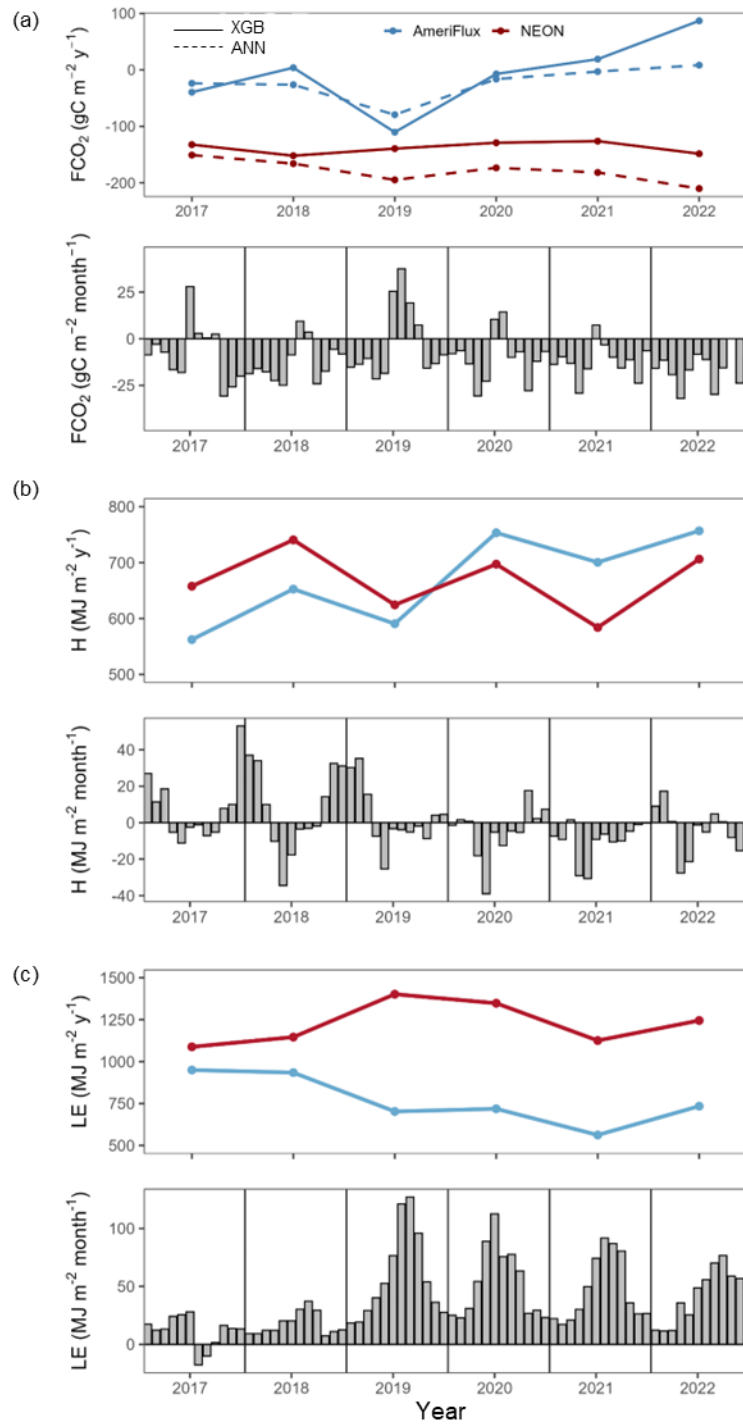


Figure 5: Annual and monthly sums of (a) carbon dioxide fluxes (FCO_2), (b) sensible heat (H), and (c) latent heat (LE) from the AmeriFlux and NEON measurements at Bartlett Experimental Forest. For each panel, the upper figure shows the annual sum, and the lower figure shows the differences between the monthly sums (NEON - AmeriFlux). For CO_2 fluxes, the annual sums of FCO_2 gap-filled using Extreme Gradient Boosting (XGBoost) are shown as solid lines, while those gap-filled using Artificial Neural Networks (ANN) are shown as dashed lines.

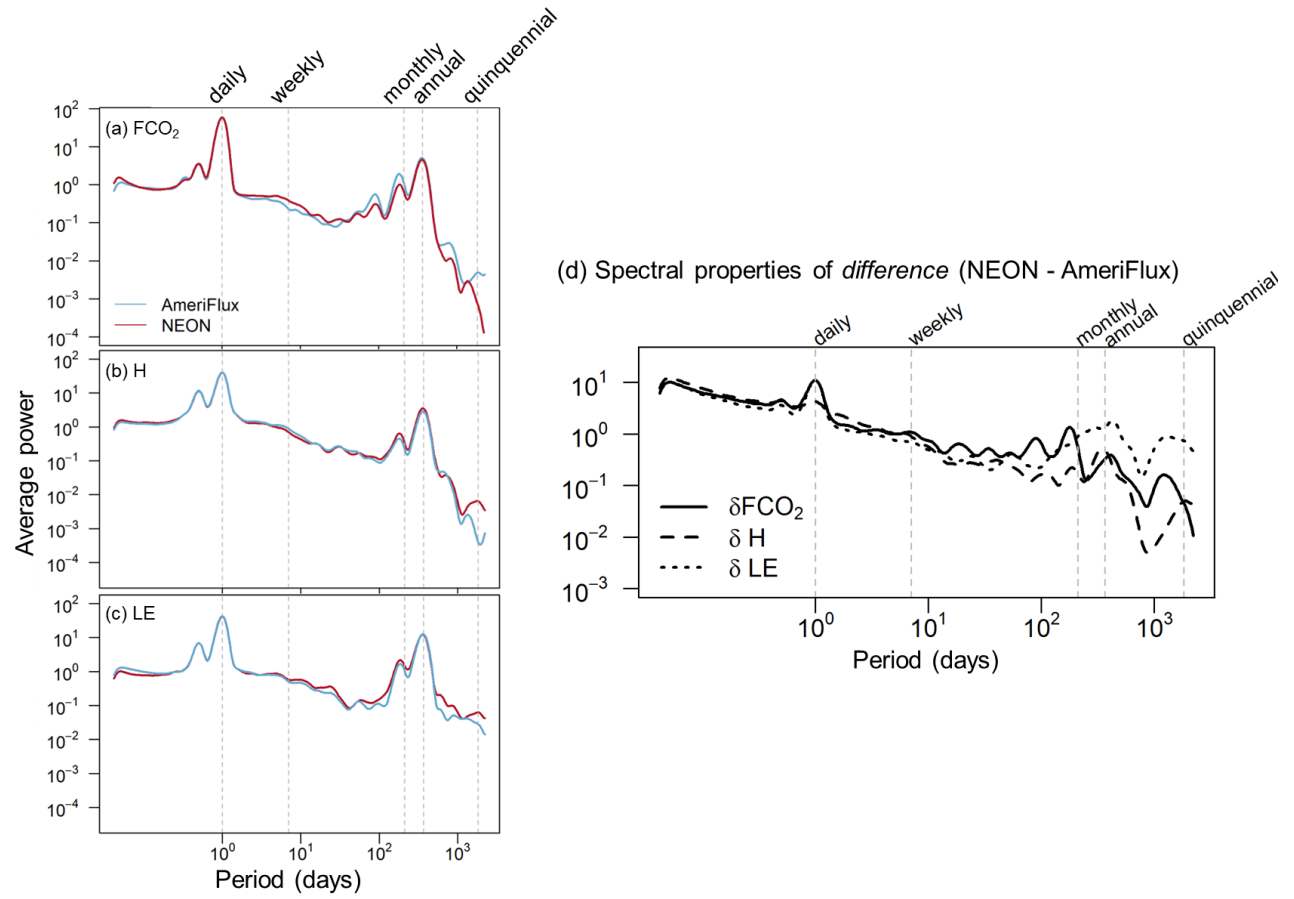


Figure 6: Wavelet power spectra (left panels) and wavelet spectra of eddy covariance time series *differences* (computed as NEON – AmeriFlux, right panel). For the left panels, from top to bottom, the variables are (a) CO₂ flux (FCO₂), (b) sensible heat flux (H), and (c) latent heat flux (LE) measured at Bartlett Experimental Forest. Wavelet power was presented in logarithmic scales (left y-axis).

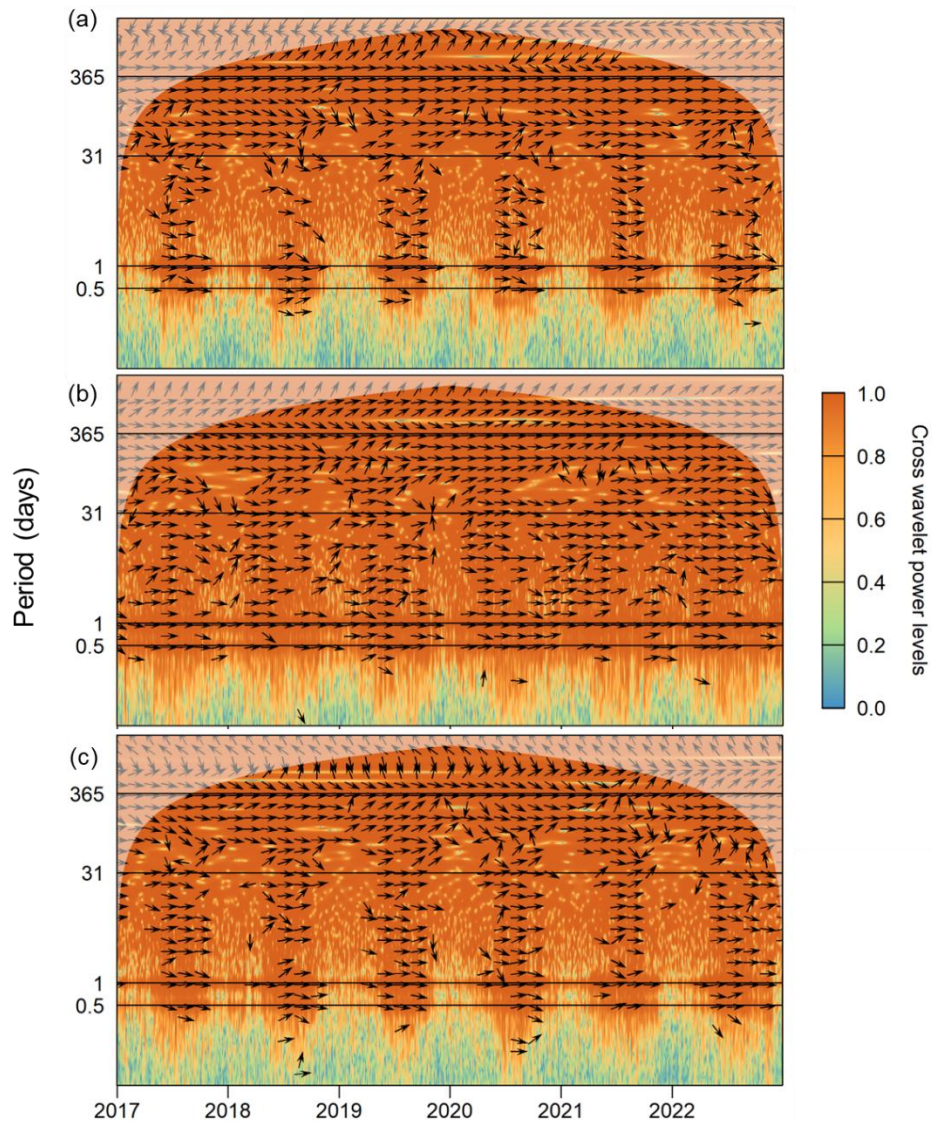


Figure 7: Scalogram of wavelet coherence between AmeriFlux and NEON measurements of (a) CO_2 flux (FCO_2), (b) sensible heat flux (H), and (c) latent heat flux (LE) measured from 2017 to 2022 at Bartlett Experimental Forest. The line that contours the semi-transparent region represents the “cone of influence”, which denotes the area beyond which wavelet coefficients are considered unreliable due to edge effects. Arrows indicate the phase with right-hand arrows indicating observations are in phase and downward arrows indicating that NEON observations lead AmeriFlux observations (and vice versa).

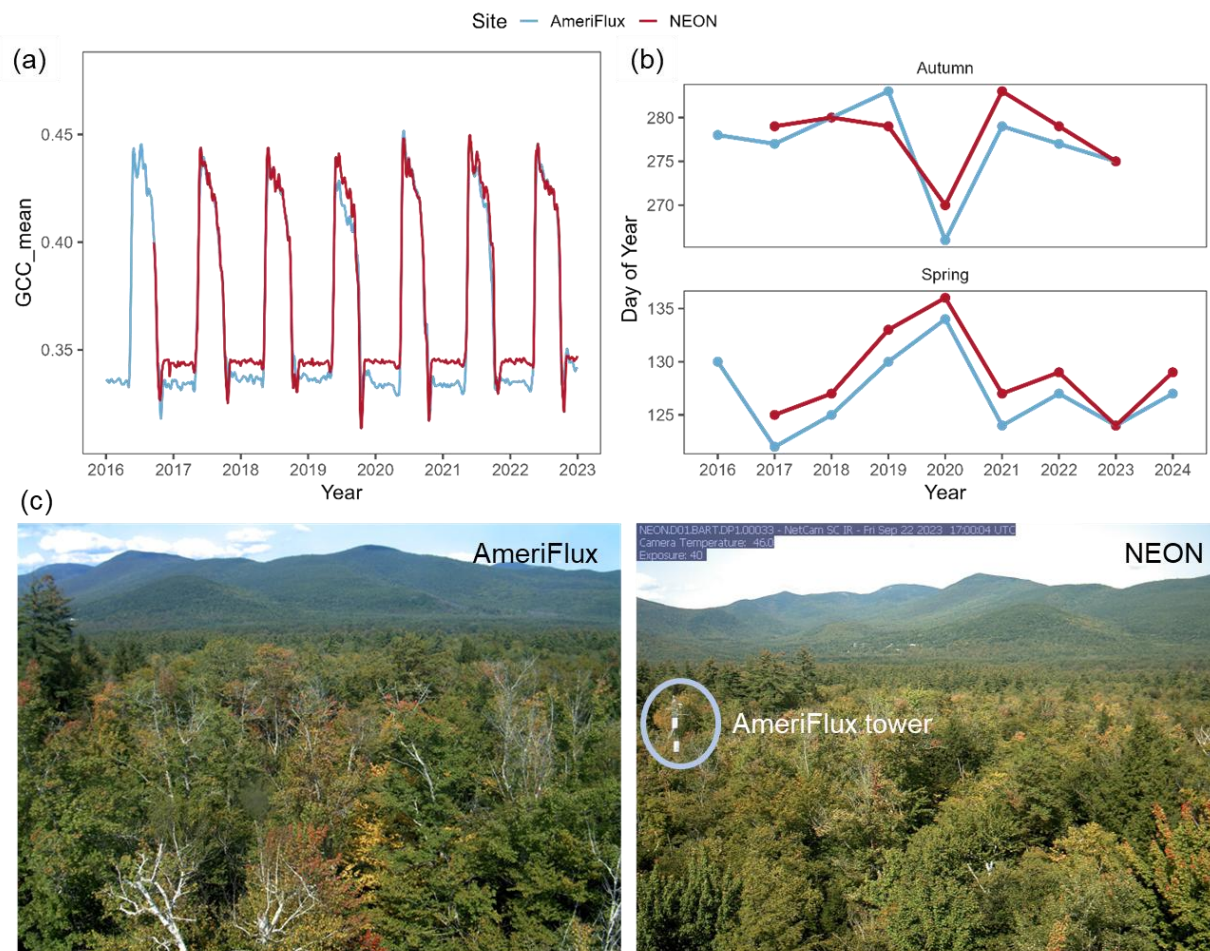


Figure 8: (a) Seasonality of green chromatic coordinate (GCC); (b) autumn and spring transition dates, represented using “transition_10” (when greenness curve hits 10% of its range, both in spring and autumn); (c) PhenoCam imagery: field of view of phenocams mounted at AmeriFlux and NEON towers. The data presented are from 3-day products.

Supporting information

Text S1. Procedures for running XGBoost and ANN.

XGBoost gap-filling procedure: For XGBoost, there are several hyperparameters that can be tuned to improve model performance, including the learning rate, depth of the trees, and regularization terms to avoid overfitting (Chen and Guestrin, 2016). To enhance computational efficiency, we used *BayesSearchCV* from *scikit-optimize* (Head et al., 2018) to explore the most promising hyperparameters within a large search space encompassing various possible values. Gaussian processes were used as surrogate models in Bayesian optimization. Early stopping was implemented, and regularization terms were fine-tuned to mitigate overfitting (Zhang and Yu, 2005). For the first step, the hyperparameters were tuned with 10-fold cross validation using the entire dataset from 2017 to 2022. Then, we evaluated model performance using 10-fold cross validation. Specifically, we randomly shuffled the entire dataset into $k=10$ folds, for use in k -fold cross validation. We trained XGB on 9 folds of data, using the 10th fold for testing. We repeated this 10 times so that every fold was included exactly once in the cross-validation set (Liu, 2024). Then the mean $\pm$ standard deviation of RMSE, R^2 , and bias on test set are reported as model performance in Table S3. In the end, we used all available data to train the model and predicted missing values in flux data.

ANN gap-filling procedure: Gap filling of FCO₂ was also performed via an artificial neural network (ANN) approach (Anderson et al., 2016; Baldocchi et al., 2015; Knox et al., 2015), using MATLAB “Deep Learning Toolbox”. Daytime and nighttime FCO₂ were gap-filled separately to ensure that performance of the ANN was not overly weighted toward more prevalent and variable daytime data. Daytime predictors included air temperature, soil temperature, VPD, global and diffuse PPFD, GCC, and time and season variables. Nighttime predictors were the same as daytime predictors excluding VPD and PPFD variables.

Uncertainties in gap-filled FCO₂ were generated by rerunning the ANN procedure twenty times with resampled training, test, and validation datasets. The median prediction was chosen as the gap-filled value and the variance as the uncertainty. For cumulative gap-filling uncertainty, the half-hourly ANN predictions for each training were summed over each year, with the variance of the annual sums representing the uncertainty.

References

- Anderson, F.E., Bergamaschi, B., Sturtevant, C., Knox, S., Hastings, L., Windham-Myers, L., Detto, M., Hestir, E.L., Drexler, J., Miller, R.L., Matthes, J.H., Verfaillie, J., Baldocchi, D., Snyder, R.L., Fujii, R., 2016. Variation of energy and carbon fluxes from a restored temperate freshwater wetland and implications for carbon market verification protocols. *JGR Biogeosciences* 121, 777–795. <https://doi.org/10.1002/2015JG003083>
- Baldocchi, D., Sturtevant, C., Contributors, F., 2015. Does day and night sampling reduce spurious correlation between canopy photosynthesis and ecosystem respiration? *Agricultural and Forest Meteorology* 207, 117–126. <https://doi.org/10.1016/j.agrformet.2015.03.010>
- Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Presented at the KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco California USA, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>
- Head, T., MechCoder, Louppe, G., Iaroslav Shcherbatyi, Fcharras, Zé Vinicius, Cmmalone, Schröder, C., Nel215, Campos, N., Young, T., Cereda, S., Fan, T., Rene-Rex, Kejia (KJ) Shi, Schwabedal, J., Carlosdanielcsantos, Hvass-Labs, Pak, M., SoManyUsernamesTaken, Callaway, F., Estève, L., Besson, L., Cherti, M., Karlson Pfannschmidt, Linzberger, F., Cauet, C., Gut, A., Mueller, A., Fabisch, A., 2018. *scikit-optimize/scikit-optimize*: v0.5.2. <https://doi.org/10.5281/ZENODO.1207017>
- Knox, S.H., Sturtevant, C., Matthes, J.H., Koteen, L., Verfaillie, J., Baldocchi, D., 2015. Agricultural peatland restoration: effects of land-use change on greenhouse gas (CO₂ and CH₄) fluxes in the Sacramento-San Joaquin Delta. *Global Change Biology* 21, 750–765. <https://doi.org/10.1111/gcb.12745>
- Liu, Y., 2024. Robust filling of extra-long gaps in eddy covariance CO₂ flux measurements from a temperate deciduous forest using eXtreme Gradient Boosting. <https://doi.org/10.5281/ZENODO.14027040>
- Zhang, T., Yu, B., 2005. Boosting with early stopping: Convergence and consistency. *Ann. Statist.* 33. <https://doi.org/10.1214/009053605000000255>

Text S2: Analysis of differences in AmeriFlux and NEON measurements for different cases.

To investigate the causes of the apparently stronger carbon sink measured by the NEON tower, despite what appears to be good agreement ($R^2 = 0.828$; [Figure 4a](#)) in the half-hour data, we divided the FCO₂ time series into four cases, where: 1) both AmeriFlux and NEON towers had valid measurements; 2) AmeriFlux had a valid measurement but NEON was filled; 3) NEON had a valid measurement but AmeriFlux was filled; 4) neither tower had a valid measurement and hence both were filled. We tallied the number of observations in each case (N_{case}), calculated the mean flux difference, ΔF ($\mu\text{mol m}^{-2} \text{s}^{-1}$) for each of these cases, and converted units to obtain the *mean* difference in C balance per 1000 measurements, $\Delta C_{\text{tot}/1000}$ (g C m^{-2}). Then we calculated the *total* difference in C balance, ΔC_{tot} (g C m^{-2}), for each case as $\Delta C_{\text{tot}/1000} \times N_{\text{case}} / 1000$. As shown in [Table S4](#), there were small differences in the measured flux for case 1 ($\Delta F = -0.28 \mu\text{mol m}^{-2} \text{s}^{-1}$), which translated to a difference of $\Delta C_{1000} = -6 \text{ g C m}^{-2}$ per 1000 half-hours. Over the 17,195 half-hours classified as case 1 in the six-year record, this accounted for a difference of $\Delta C_{\text{tot}} = -104 \text{ g C m}^{-2}$ (or, on an annual basis, $-17 \text{ g C m}^{-2} \text{y}^{-1}$) where the negative value corresponds to more negative annual FCO₂ for the NEON data. Differences in the underlying measured values then propagated, in a similar manner, to the respective filled values for cases 2-4, further enhancing the stronger sink indicated by the NEON tower flux measurements. Notably, cases 3 and 4 together accounted for a difference of $-93 \text{ g C m}^{-2} \text{y}^{-1}$, or more than two-thirds of the stronger NEON C sink.

Dividing ([Table S3](#)) the data record into nighttime (PPFD $< 5 \mu\text{mol m}^{-2} \text{s}^{-1}$) and daytime (PPFD $\geq 5 \mu\text{mol m}^{-2} \text{s}^{-1}$) half-hours, we found that for each of the four cases at night, the data again indicate a stronger NEON C sink. However, across all four cases, nighttime differences accounted for only $\Delta C_{\text{tot}} = -217 \text{ g C m}^{-2}$ over the 6-year record, or $-36 \text{ g C m}^{-2} \text{y}^{-1}$. Therefore, although differences in measurements propagated to the respective filled values for cases 2-4, nighttime data still accounted for only about one-quarter of the $130 \text{ g C m}^{-2} \text{y}^{-1}$ difference in annual sink strength. By contrast, ΔF was much larger in magnitude during the day for each of the 4 cases, which then resulted in substantial differences in ΔC_{tot} ($\Delta C_{\text{tot}} = -563 \text{ g C m}^{-2}$ over the 6-year record, or $-94 \text{ g C m}^{-2} \text{y}^{-1}$), particularly for case 3 (-220 g C m^{-2} across the six-year record, or $-37 \text{ g C m}^{-2} \text{y}^{-1}$).

Breaking the daytime cases down by month, we found strong seasonal patterns in the frequency of occurrence of case 1, with far more instances where both AmeriFlux and NEON towers had valid measurements ([Supplemental Figure S15](#)). Case 2 occurred more commonly in the autumn months, whereas case 3 occurred more commonly in the spring months. There was no clear pattern in the occurrence of case 4 across months. For all 4 cases, however, there was strong seasonality in ΔF and hence $\Delta C_{\text{tot}/1000}$ ([Figure S15](#)), which were then amplified by the seasonality of $N_{\text{cases}/\text{mo}}$. While June and July uptake was marginally stronger for the AmeriFlux tower (positive values, especially for case 2), spring (April and May, especially cases 1 and 3) and Autumn (October and November, especially cases 1, 2, and 3) uptake was substantially stronger for the NEON tower. Combined with the frequency of occurrence of each case, April and May (aggregate $\Delta C_{\text{tot}} = -263 \text{ g C m}^{-2}$ over the six-year record, or $-44 \text{ g C m}^{-2} \text{y}^{-1}$) and October and November (aggregate $\Delta C_{\text{tot}} = -219 \text{ g C m}^{-2}$ over the six year record, or $-36 \text{ g C m}^{-2} \text{y}^{-1}$) together accounted for two-thirds of the overall difference in sink strength between the two towers. The seasonality of these patterns is consistent with the NEON tower footprint including a higher proportion of evergreen vegetation ([Figure 3](#), [Figure S10](#)). Evergreen conifers are photosynthetically more active during these “shoulder season” months than the deciduous trees, which does not develop a full canopy until late May, and which begins to turn color in mid-September.

Table S1. Sensors and measurement protocols of US-Bar (AmeriFlux) and US-xBR (NEON) at Bartlett Experiment Forest (Bartlett Experimental Forest). Measurement heights for sensors used in this analysis are reported in [Table 1](#).

SiteID	US-Bar (AmeriFlux)	US-xBR (NEON)
Tower height	26.5 m	35.7 m
Flux measurement height	24.5 m	35.7 m
Sonic anemometer	SAT-211/ 3 K 3- axis sonic anemometer (Applied Technologies, Longmont, Colo). Data is collected at 5 Hz.	CSAT-3 sonic anemometer (Campbell Scientific Inc., Logan, Utah, USA). Data is collected at 20 Hz. Wind speed and wind direction are measured at 6 different heights: 35.7, 25.5, 19.2, 15.5, 2.48 and 0.56 m.
Infrared gas analyzer (IRGA)	LI-6262 CO ₂ /H ₂ O infrared gas analyzer (Li-Cor, Lincoln NE, USA). In 2014 the LI-6262 was replaced with a model LI-7200 analyzer (Oumette et al., 2018). Data is recorded at 5 Hz.	An enclosed path gas analyzer of LI-7200 (Li-Cor Inc., Lincoln, Nebraska, USA) records CO ₂ and H ₂ O concentration at 20 Hz.
Storage correction	No	LI840A and LI850 gas analyzers record the CO ₂ and H ₂ O concentration at 1 Hz.
Raw data processing	EddyPro	Eddy4R
Quality control protocols	Filtering for low wind speed, sonic anemometer w and T variance (high and low thresholds), IRGA H ₂ O and CO ₂ variance (high and low thresholds) (Richardson et al., 2019)	
Correction of high pass filtering effects	Following Moncrieff et al. (2008)	Yes, simplified version of adaptive/semi-empirical Nordbo and Katul, 2012
Correction of low pass filtering	Yes, following Ibrom et al. (2007)	No
Time lags compensation	Covariance maximization with default	Adaptive maximum cross correlation after high-pass filtering (Hartmann et al., 2018)
Air Temperature	Platinum resistance thermometer, with thermistor (HMP35 and HMP155 humidity and temperature probe, Vaisala Oyj, Helsinki, Finland) backup, sampled every 10 s with half-hourly averages logged.	The air temperature profile is measured with aspirated temperature sensors (MetOne Instruments, Inc., model 076B-7388; Grant Pass, Oregon, USA). The measurements are taken at 6 different heights: 35.2, 35.0, 25.5, 19.2, 15.5, 2.5, and 0.1 m. Sensors are sampled at 1 Hz and one- and thirty-mins averages are logged.
Total, Diffuse, and direct Photosynthetic photon flux density (PPFD) of photosynthetically active radiation (PAR)	Quantum sensor (LI-190, Li-Cor Environmental, Lincoln, NE, USA) for total PPFD. BF5 Sunshine Sensor (Delta-T Devices Ltd., Cambridge, UK) for direct and diffuse PPFD. Sensors sampled every 10 s with half-hourly averages logged.	Pyranometer (CMP22, Kipp & Zonen, Delft, The Netherlands) and sunshine pyranometer (SPN1, Delta-T Devices Ltd., Cambridge, UK. Measurements are made by sensors located at 6 different heights: 35.6, 26.1, 19.8, 16.1, 3.11 and 0.73. Sensors are sampled at 1 Hz and half-hourly averages are logged.
Net radiation, net shortwave radiation, net longwave radiation	2-channel net radiometer (CNR2, Kipp & Zonen, Delft, The Netherlands) measures net shortwave and net longwave components, which are summed to obtain net radiation. Sensor sampled every 10 s with half-hourly averages logged.	Observations are sampled at 1 Hz and provided as one- and thirty-minute averages. Sensors (Hukseflux NR01 Net Radiometer) are located at the top of the tower infrastructure (for shortwave and longwave). Shortwave up and down are measured using CM3 pyranometer from Kipp and Zonen.
Relative humidity	HMP35 and HMP155 humidity and temperature probe, Vaisala Oyj, Helsinki, Finland). Sensors sampled	Vaisala HUMICAP humidity and temperature probes - HMP 155. Sensors

	every 10 s with half-hourly averages logged.	are sampled at 1 Hz and one- and thirty-mins averages are logged.
Soil temperature	T-type thermocouple. Sensors sampled every 10 s with half-hourly averages logged.	Soil temperature is measured using platinum resistance thermometers installed at various depths in the soil. At Bartlett Experimental Forest, there are 45 sensors positioned at 20 different depths, ranging from 0.03 to 1.97 m. Sensors are sampled at 10 Hz and half-hourly averages are logged.
Soil volumetric water content	Time-domain reflectometer (CS616, Campbell Scientific, Logan UT, USA)	Data are measured using capacitance sensors and are collected from all five instrumented soil plots at each site.
Soil heat flux plate	Not available	Sensors were initially installed at ~8 cm deep at most sites and are in the process of being moved to ~5 cm deep.
PhenoCam	StarDot NetCam SC IR 5 MP (StarDot Technologies, Buena Park CA, USA)	StarDot NetCam SC IR 5 MP (StarDot Technologies, Buena Park CA, USA)

References:

- Ibrom, A., Dellwik, E., Flyvbjerg, H., Jensen, N.O., Pilegaard, K., 2007. Strong low-pass filtering effects on water vapour flux measurements with closed-path eddy correlation systems. *Agricultural and Forest Meteorology* 147, 140–156. <https://doi.org/10.1016/j.agrformet.2007.07.007>
- Moncrieff, J., Clement, R., Finnigan, J., Meyers, T., 2005. Averaging, detrending, and filtering of Eddy Covariance time series, in: Lee, X., Massman, W., Law, B. (Eds.), *Handbook of Micrometeorology*, Atmospheric and Oceanographic Sciences Library. Kluwer Academic Publishers, Dordrecht, pp. 7–31. https://doi.org/10.1007/1-4020-2265-4_2
- Nordbo, A., Katul, G., 2013. A wavelet-based correction method for Eddy-Covariance high-frequency losses in scalar concentration measurements. *Boundary-Layer Meteorol* 146, 81–102. <https://doi.org/10.1007/s10546-012-9759-9>
- Ouimette, A.P., Ollinger, S.V., Richardson, A.D., Hollinger, D.Y., Keenan, T.F., Lepine, L.C., Vadeboncoeur, M.A., 2018. Carbon fluxes and interannual drivers in a temperate forest ecosystem assessed through comparison of top-down and bottom-up approaches. *Agricultural and Forest Meteorology* 256–257, 420–430. <https://doi.org/10.1016/j.agrformet.2018.03.017>
- Richardson, A.D., Hollinger, D.Y., Shoemaker, J.K., Hughes, H., Savage, K., Davidson, E.A., 2019. Six years of ecosystem-atmosphere greenhouse gas fluxes measured in a sub-boreal forest. *Sci Data* 6, 117. <https://doi.org/10.1038/s41597-019-0119-1>

Table S2. Data products used in this paper as leaf and canopy traits.

Name and unit	Description	Resolution
Canopy nitrogen map (%)	The canopy nitrogen map (University Of New Hampshire Earth Systems Research Center, 2016) was created from airborne imaging spectrometer data created using data from NASA's AVIRIS (Airborne Visible/InfraRed Imaging Spectrometer) instrument, as described in Ollinger et al. (2008).	15.9 m
Canopy height model (m)	The canopy height model (data product DP3.30015.001; NEON, 2024), was derived from NEON Airborne Observation Platform data (AOP, NEON, 2023b), where the canopy height model is derived directly from a LiDAR point cloud.	1 m
Enhanced Vegetation Index (EVI_{max} , unitless)	The map of EVI_{max} (unitless) was derived from PlanetScope imagery (Moon et al., 2022).	3 m
Vegetation area index (VAI, unit: m^2/m^2)	The VAI map (data product DP1.30003.001; NEON, 2024) was derived from discrete return LiDAR point cloud data from a NEON AOP flight in 2022, with a cut-off height of 0.5 m and grain size of 10 m (Wang et al., 2024).	30 m

References:

- Moon, M., Richardson, A.D., Milliman, T., Friedl, M.A., 2022. A high spatial resolution land surface phenology dataset for AmeriFlux and NEON sites. *Sci Data* 9, 448. <https://doi.org/10.1038/s41597-022-01570-5>
- Ollinger, S.V., Richardson, A.D., Martin, M.E., Hollinger, D.Y., Frohking, S.E., Reich, P.B., Plourde, L.C., Katul, G.G., Munger, J.W., Oren, R., Smith, M.-L., Paw U, K.T., Bolstad, P.V., Cook, B.D., Day, M.C., Martin, T.A., Monson, R.K., Schmid, H.P., 2008. Canopy nitrogen, carbon assimilation, and albedo in temperate and boreal forests: Functional relations and potential climate feedbacks. *Proc. Natl. Acad. Sci. U.S.A.* 105, 19336–19341. <https://doi.org/10.1073/pnas.0810021105>
- University Of New Hampshire Earth Systems Research Center, 2016. Canopy Nitrogen, Bartlett Experimental Forest, NH USA, 2012. <https://doi.org/10.5063/F1V9860T>
- Wang, J., Choi, D.H., LaRue, E.A., Atkins, J.W., Foster, J.R., Matthes, J.H., Fahey, R.T., Fei, S., Hardiman, B.S., 2024. Neon-Sd: A 30-M Structural Diversity Product Derived from the Neon Discrete-Return Lidar Point Cloud. <https://doi.org/10.2139/ssrn.4848574>

Table S3: Performance of XGBoost in gap-filling flux data: CO₂ flux (FCO₂), sensible heat flux (H), and latent heat flux (LE) measured from 2017 to 2022 at Bartlett Experimental Forest. Based on cross validation, mean $\pm$ standard deviation of RMSE, R² and bias on test set are reported.

	RMSE	R ²	Bias
AmeriFlux FCO ₂	2.60 $\pm$ 0.06 $\mu\text{mol m}^{-2} \text{s}^{-1}$	0.879 $\pm$ 0.004	-0.00 $\pm$ 0.04 $\mu\text{mol m}^{-2} \text{s}^{-1}$
- Daytime	2.64 $\pm$ 0.07 $\mu\text{mol m}^{-2} \text{s}^{-1}$	0.873 $\pm$ 0.005	-0.00 $\pm$ 0.04 $\mu\text{mol m}^{-2} \text{s}^{-1}$
- Nighttime	2.48 $\pm$ 0.12 $\mu\text{mol m}^{-2} \text{s}^{-1}$	0.394 $\pm$ 0.051	-0.02 $\pm$ 0.14 $\mu\text{mol m}^{-2} \text{s}^{-1}$
NEON FCO ₂	2.49 $\pm$ 0.04 $\mu\text{mol m}^{-2} \text{s}^{-1}$	0.873 $\pm$ 0.004	-0.00 $\pm$ 0.02 $\mu\text{mol m}^{-2} \text{s}^{-1}$
- Daytime	2.73 $\pm$ 0.04 $\mu\text{mol m}^{-2} \text{s}^{-1}$	0.852 $\pm$ 0.005	0.00 $\pm$ 0.03 $\mu\text{mol m}^{-2} \text{s}^{-1}$
- Nighttime	1.53 $\pm$ 0.06 $\mu\text{mol m}^{-2} \text{s}^{-1}$	0.638 $\pm$ 0.035	-0.02 $\pm$ 0.05 $\mu\text{mol m}^{-2} \text{s}^{-1}$
AmeriFlux H	39.90 $\pm$ 0.68 W m ⁻²	0.857 $\pm$ 0.005	-0.17 $\pm$ 0.50 W m ⁻²
NEON H	33.29 $\pm$ 0.689 W m ⁻²	0.888 $\pm$ 0.005	-0.17 $\pm$ 0.40 W m ⁻²
AmeriFlux LE	18.80 $\pm$ 0.26 W m ⁻²	0.912 $\pm$ 0.004	-0.00 $\pm$ 0.31 W m ⁻²
NEON LE	27.10 $\pm$ 0.59 W m ⁻²	0.891 $\pm$ 0.003	-0.02 $\pm$ 0.57 W m ⁻²

Table S4. Difference in measured and gapfilled FCO₂ measurements from NEON and AmeriFlux towers.

Case	All data					Nighttime Data (PPFD < 5 $\mu\text{mol m}^{-2} \text{s}^{-1}$)					Daytime Data (PPFD $\geq$ 5 $\mu\text{mol m}^{-2} \text{s}^{-1}$)				
	N _{case}	ΔF	$\Delta C_{\text{tot}/1000}$	ΔC_{tot}	$\Delta C_{\text{tot}/y}$	N _{case}	ΔF	$\Delta C_{\text{tot}/1000}$	ΔC_{tot}	$\Delta C_{\text{tot}/y}$	N _{case}	ΔF	$\Delta C_{\text{tot}/1000}$	ΔC_{tot}	$\Delta C_{\text{tot}/y}$
1	17,195	-0.28	-6.1	-104	-17	2,891	-0.09	-2.0	-6	-1	14,304	-0.32	-6.9	-98	-16
2	13,286	-0.42	-9.0	-120	-20	4,052	-0.12	-2.6	-10	-2	9,234	-0.55	-11.9	-109	-18
3	19,972	-0.61	-13.2	-264	-44	6,217	-0.32	-7.0	-43	-7	13,755	-0.74	-16.0	-220	-37
4	54,715	-0.25	-5.3	-292	-49	38,938	-0.19	-4.0	-157	-26	15,777	-0.40	-8.5	-135	-22
Total	105,168	-0.34	-7.4	-780	-130	52,098	-0.19	-4.2	-217	-36	53,070	-0.49	-10.6	-563	-94

Notes: Cases are as follows: 1) both AmeriFlux and NEON towers had valid measurements; 2) AmeriFlux had a valid measurement, but NEON was filled; 3) NEON had a valid measurement but AmeriFlux was filled; 4) neither tower had a valid measurement and hence both were filled. N_{case} is the number of half hours classified as each case, across all years. ΔF is mean difference in FCO₂ (measured or gap-filled, depending on the case) in $\mu\text{mol m}^{-2} \text{s}^{-1}$. $\Delta C_{\text{tot}/1000}$ is the total difference in C uptake between the NEON and AmeriFlux towers in g C m⁻² per 1000 half-hourly measurements. ΔC_{tot} is the total overall difference in C uptake (g C m⁻²) for each case ($\Delta C_{\text{tot}/1000} \times \text{N}_{\text{case}} / 1000$), across all 6 years. $\Delta C_{\text{tot}/y}$ is the total annual difference in C uptake for each case (g C m⁻²/y, calculated as ΔC_{tot} divided by 6 years). In the total row, N_{case}, ΔC_{tot} , and $\Delta C_{\text{tot}/y}$ are summed over all cases, but ΔF and $\Delta C_{\text{tot}/1000}$ are weighted (by N_{case}) means. Negative values indicate a more negative C balance (more uptake) by the NEON tower relative to the AmeriFlux tower.

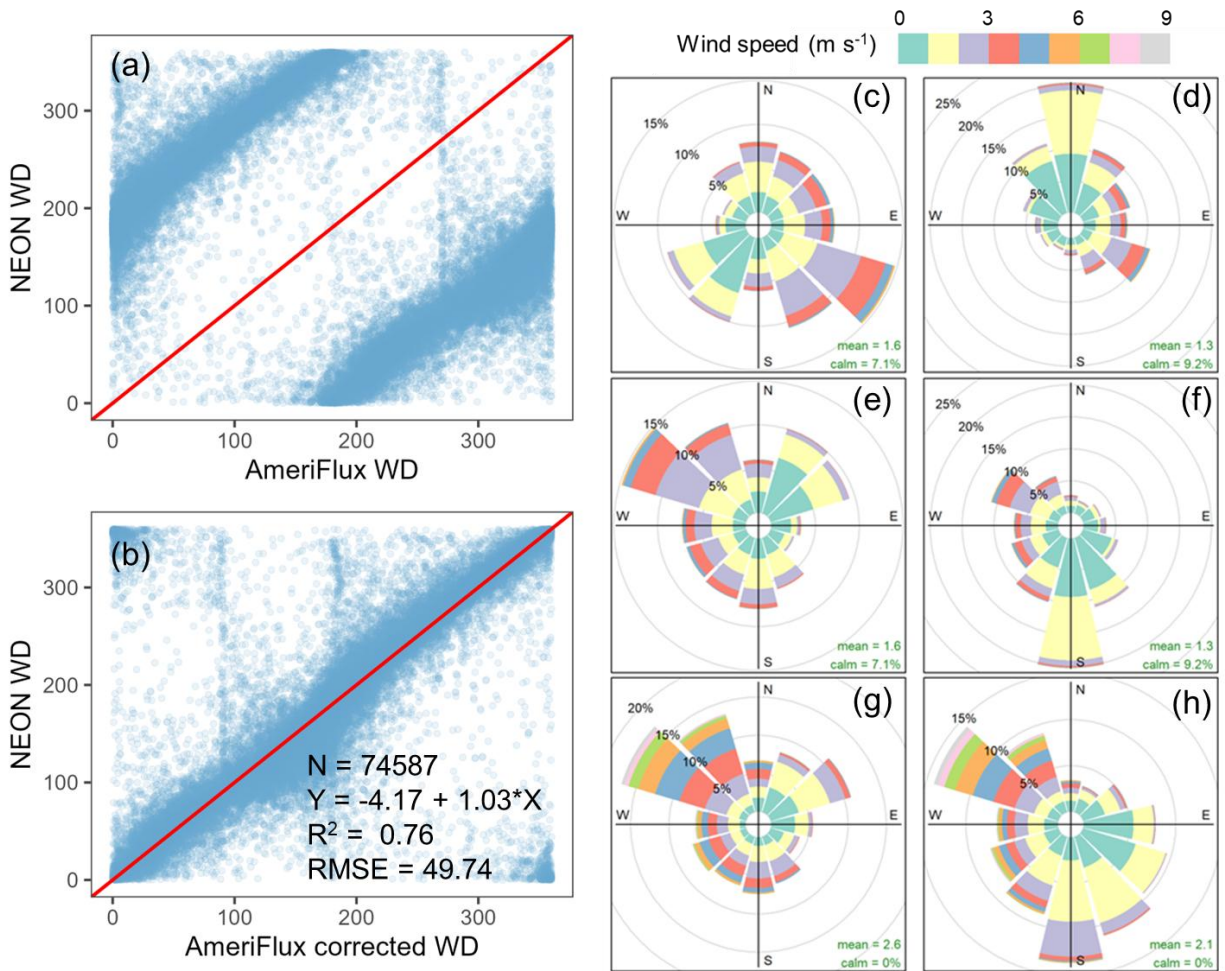


Figure S1: Wind speed and wind direction: (a) Scatter plot of raw wind direction data from AmeriFlux vs. NEON; (b) scatter plot of corrected wind direction from AmeriFlux (accounting for the sonic anemometer installed upside-down for the duration of the period of comparison) vs. NEON; (c) Daytime and (d) nighttime wind rose of AmeriFlux without correction for wind direction; (e) Daytime and (f) nighttime wind rose of AmeriFlux after correction for wind direction; (g) Daytime and (h) nighttime wind rose of NEON. The red lines in (a) and (b) represent the one-to-one line. The AmeriFlux wind direction is corrected by adding 180 degrees to the original direction and applying the modulus operation with 360.

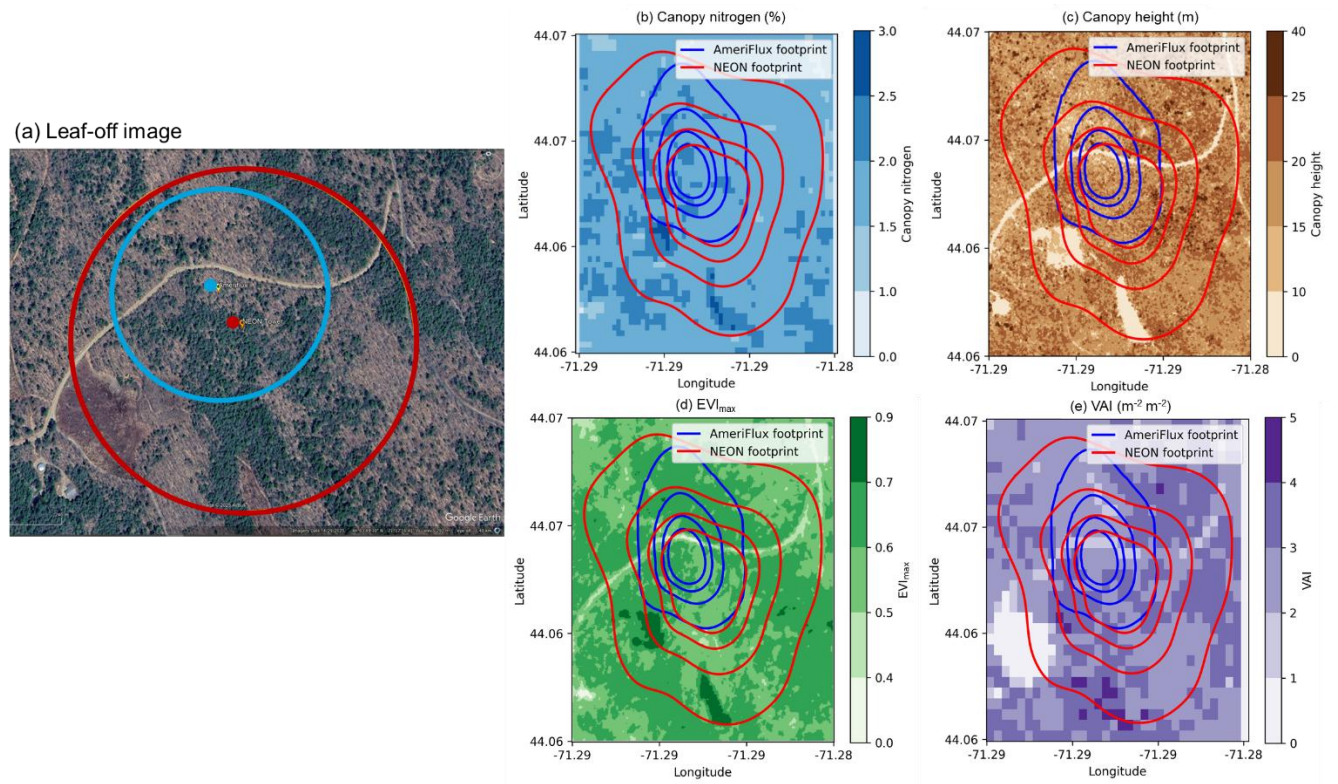


Figure S2: Daytime footprints for AmeriFlux and NEON with (a) leaf-off image from Google Earth, with blue and red circles representing the approximate footprint areas for AmeriFlux and NEON respectively. (b) canopy nitrogen (%), (c) canopy height (m), (d) EVI_{max} (unitless) and (e) vegetation area index (VAI, $m^2 m^{-2}$).

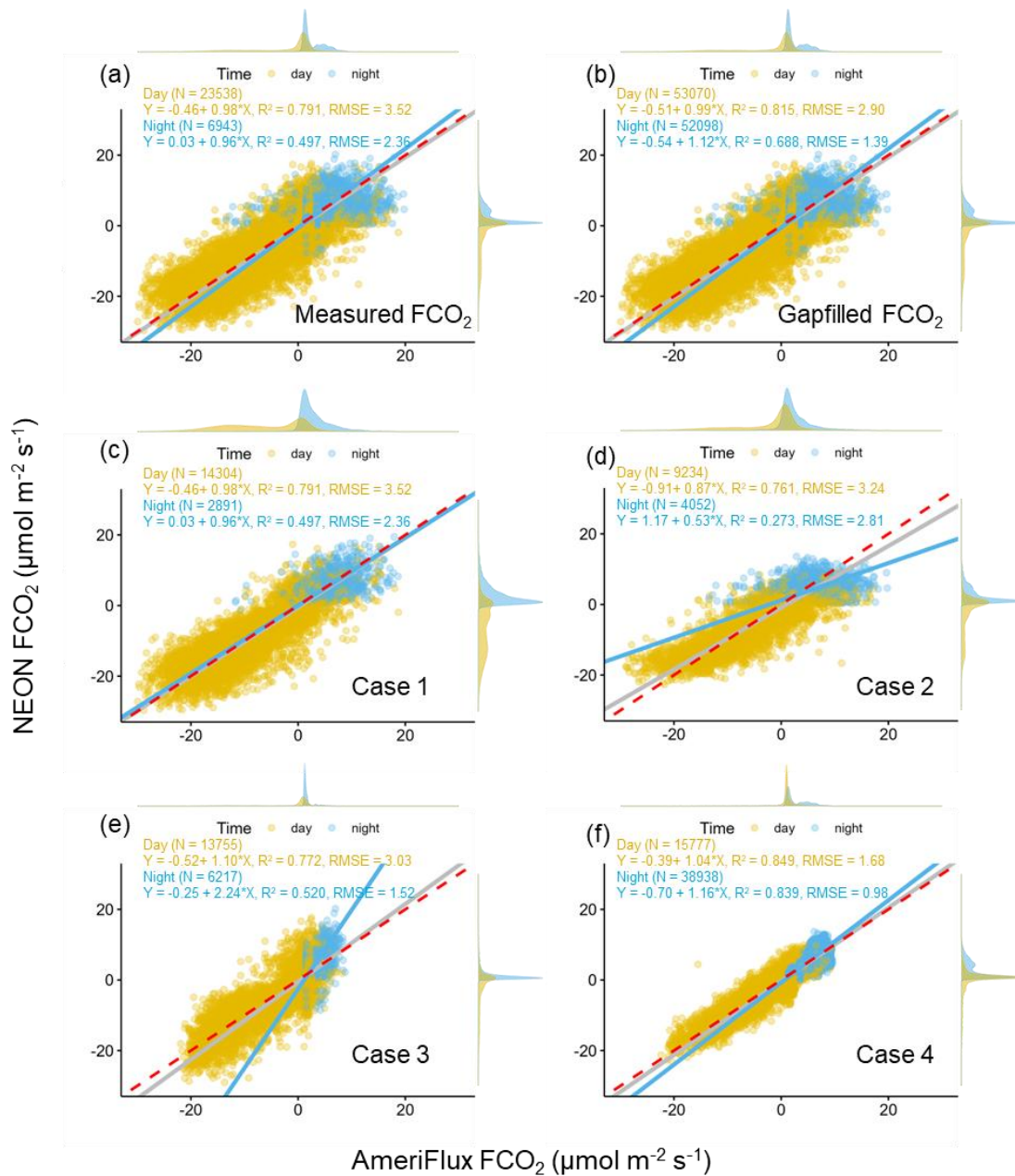


Figure S3. (a) Scatter plot comparing simultaneous measurements of half-hourly carbon dioxide fluxes (FCO₂) from AmeriFlux and NEON at daytime (yellow) and nighttime (blue), respectively; (b) Scatter plots comparing gapfilled carbon dioxide fluxes (FCO₂) from AmeriFlux and NEON at daytime (yellow) and nighttime (blue), respectively. Panels (c)–(f) represent different cases: c) both AmeriFlux and NEON towers had valid measurements; d) AmeriFlux had a valid measurement, but NEON was filled; e) NEON had a valid measurement but AmeriFlux was filled; f) neither tower had a valid measurement and hence both were filled. Daytime data are defined when incoming PPFD is greater than 5 μmol m⁻² s⁻². The R², slope, and intercept between the variables are evaluated using Model II regression.

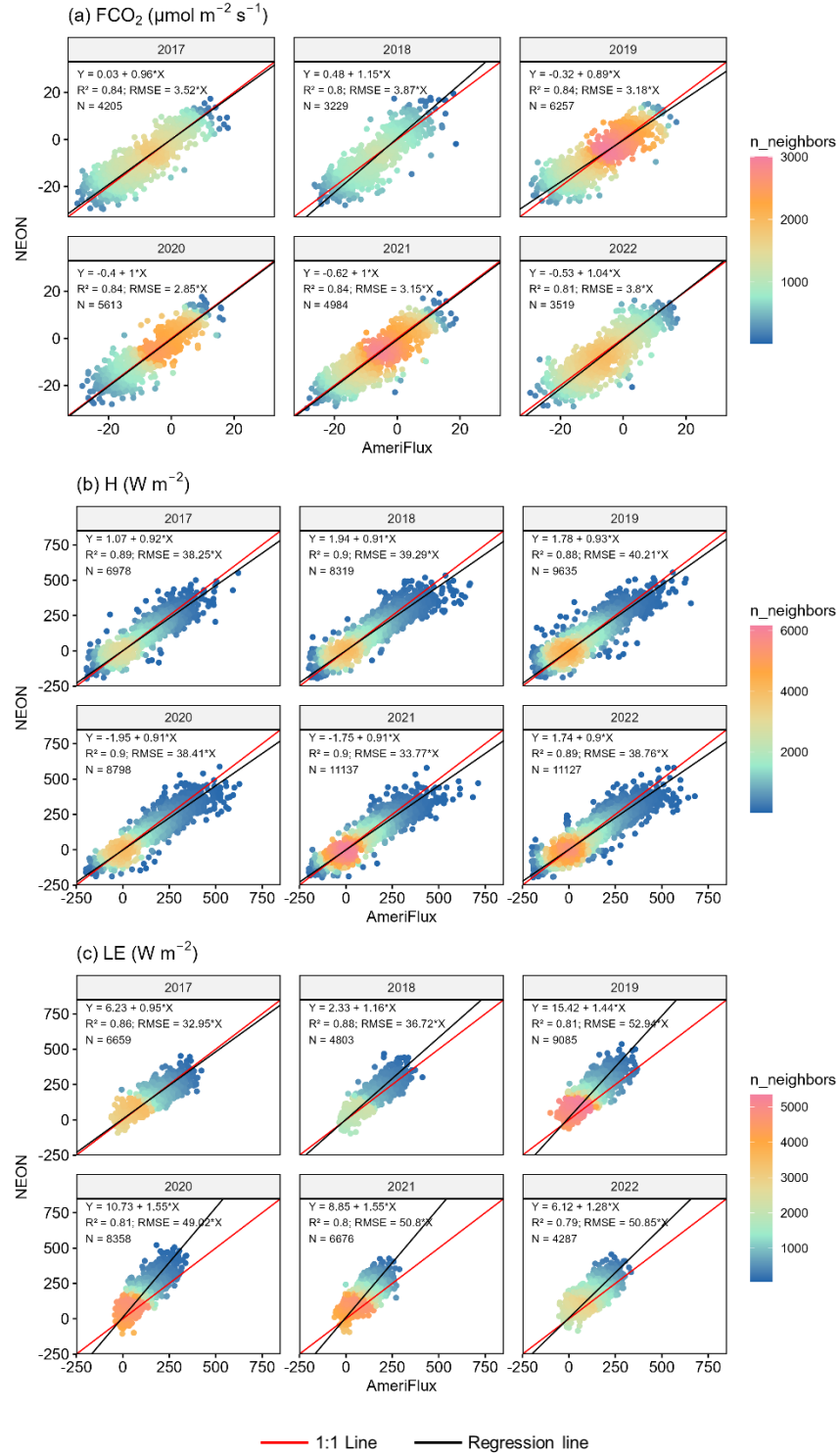


Figure S4. Scatter density plots comparing simultaneous half-hourly measurements of (a) CO_2 flux (FCO_2), (b) sensible heat flux (H), and (c) latent heat flux (LE) from AmeriFlux and NEON at Bartlett Experimental Forest for each year. Individual points are colored by the number of neighboring points ($n_neighbors$). Regression lines are from Model II regression.

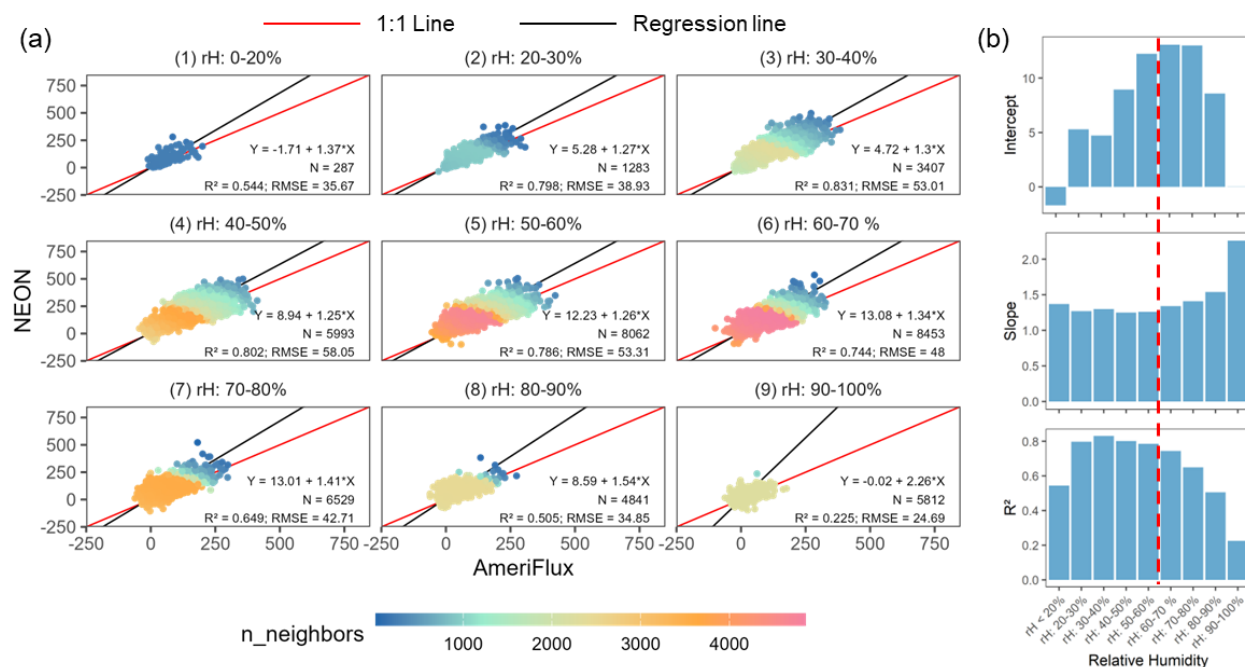


Figure S5. (a) Scatter density plot comparing simultaneous measurements of half-hourly latent heat (LE) from AmeriFlux and NEON at different relative humidity (rH) levels (LE units: W m^{-2}). Individual points are colored by the number of neighboring points ($n_{\text{neighbors}}$). (b) The intercept, slope and R^2 between LE measured by AmeriFlux and NEON evaluated using Model II regression. The red dash line here indicates when there is an obvious decrease in regression.

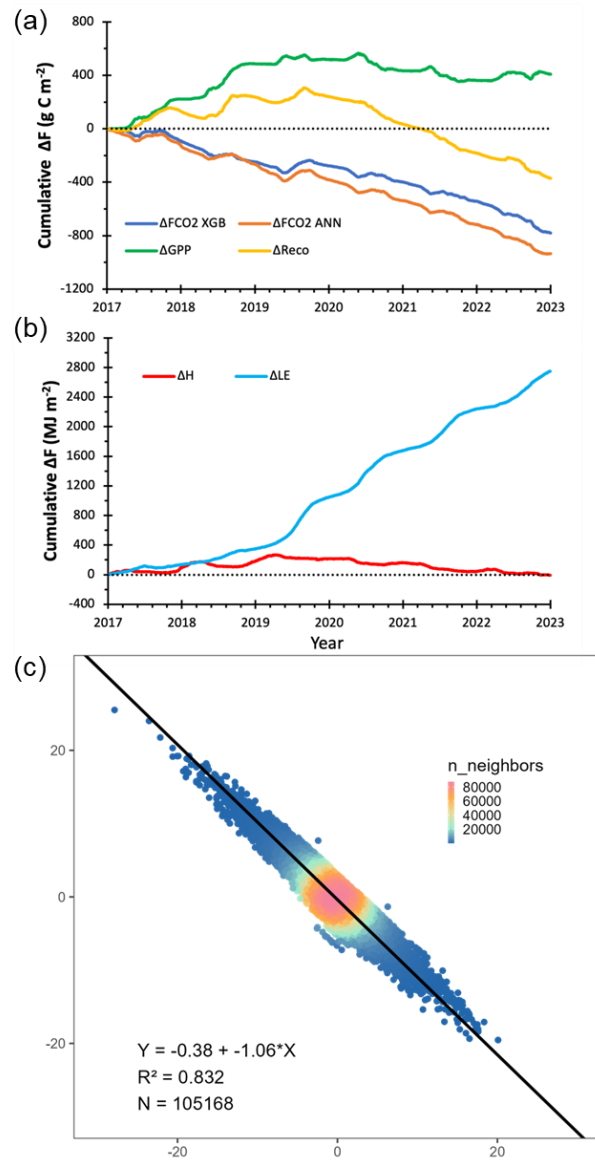


Figure S6. Cumulative differences in measured and filled fluxes over the 6-y period (2017-2022) analyzed here. Differences are calculated as NEON – AmeriFlux. (a) cumulative difference in measured and gap-filled FCO_2 (showing both XGB and ANN gap-filling), as well as the cumulative difference in partitioned estimates of both GPP and Reco; (b) cumulative difference in sensible and latent heat flux. In (a), it is notable that the evolution of ΔFCO_2 is extremely similar for both XGB and ANN gap filling, giving confidence in the temporal patterns and annual sums. Additionally, it can be seen how, because of the nighttime partitioning method we used, the cumulative ΔGPP reflects the combined effects of ΔFCO_2 and ΔReco . In (b), cumulative ΔLE accelerates beginning in 2019, reaching 2750 MJ m^{-2} by 2023 (an average difference of $460 \text{ MJ m}^{-2} \text{ y}^{-1}$) whereas by the end of 2023, the cumulative ΔH is only 6 MJ m^{-2} , or $1 \text{ MJ m}^{-2} \text{ y}^{-1}$.

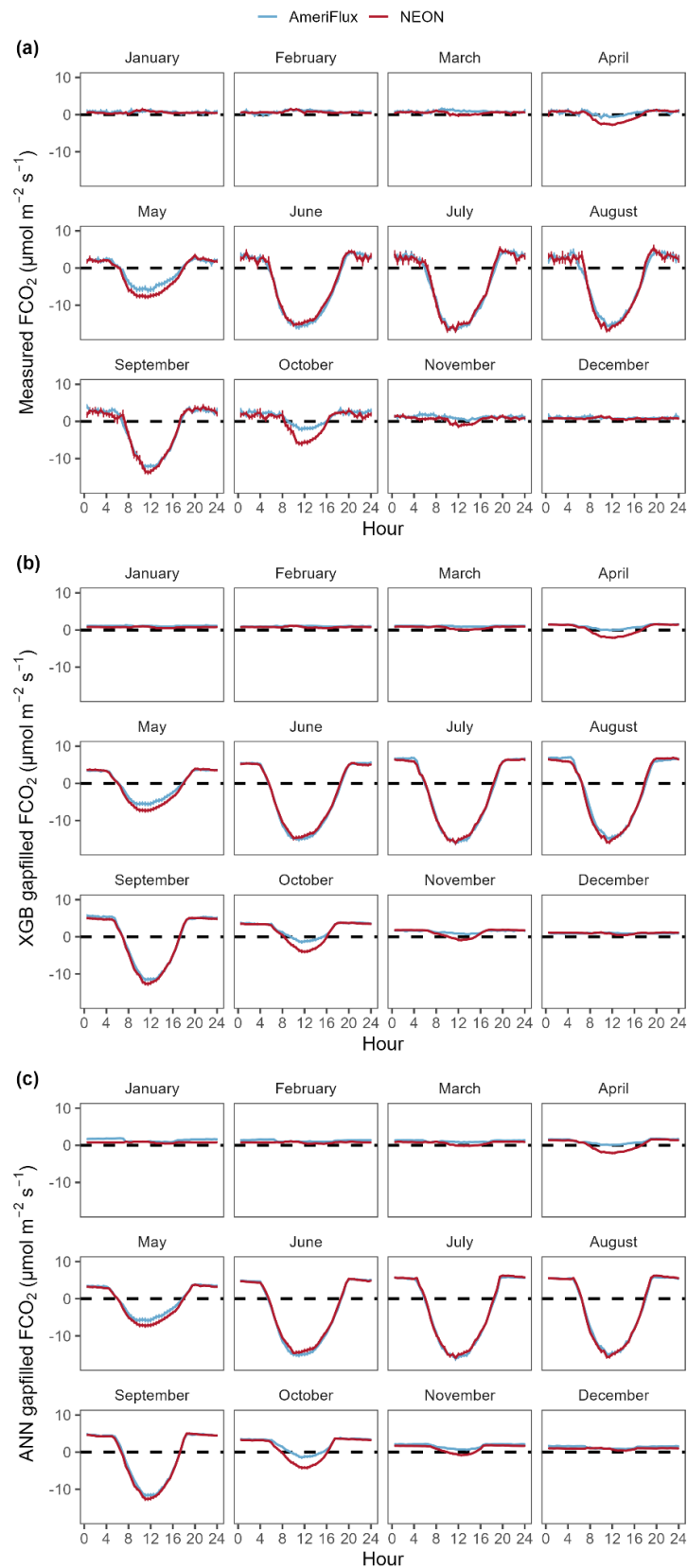


Figure S7. Average diel pattern of (a) measured CO₂ flux (FCO₂), (b) XGBoost gapfilled and FCO₂ and (c) ANN gapfilled FCO₂ by month for all years combined. Error bars show standard error within each half-hourly bin.

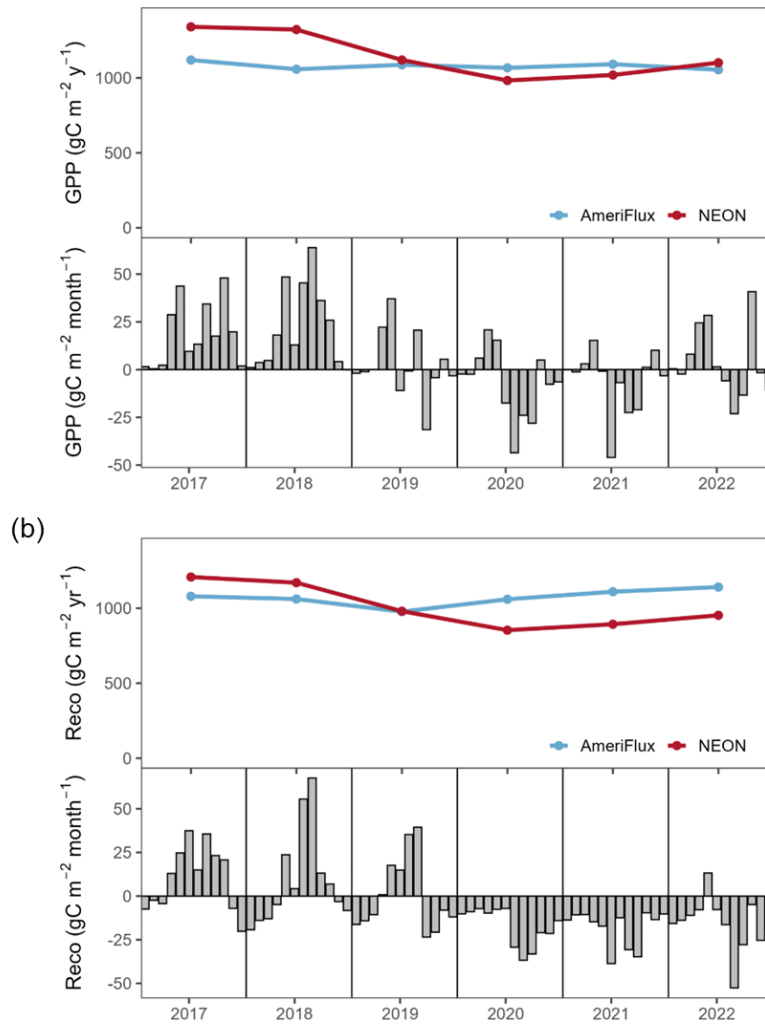


Figure S8. Annual and monthly sums of (a) gross primary production (GPP) and (b) ecosystem respiration (Reco) from AmeriFlux and NEON measurements at Bartlett Experimental Forest. For each panel, the upper figure displays the annual sums, with AmeriFlux shown as a blue line and NEON as a red line. The lower figure shows the differences in monthly sums, calculated as the difference between AmeriFlux and NEON monthly sums (NEON - AmeriFlux) for GPP or Reco.

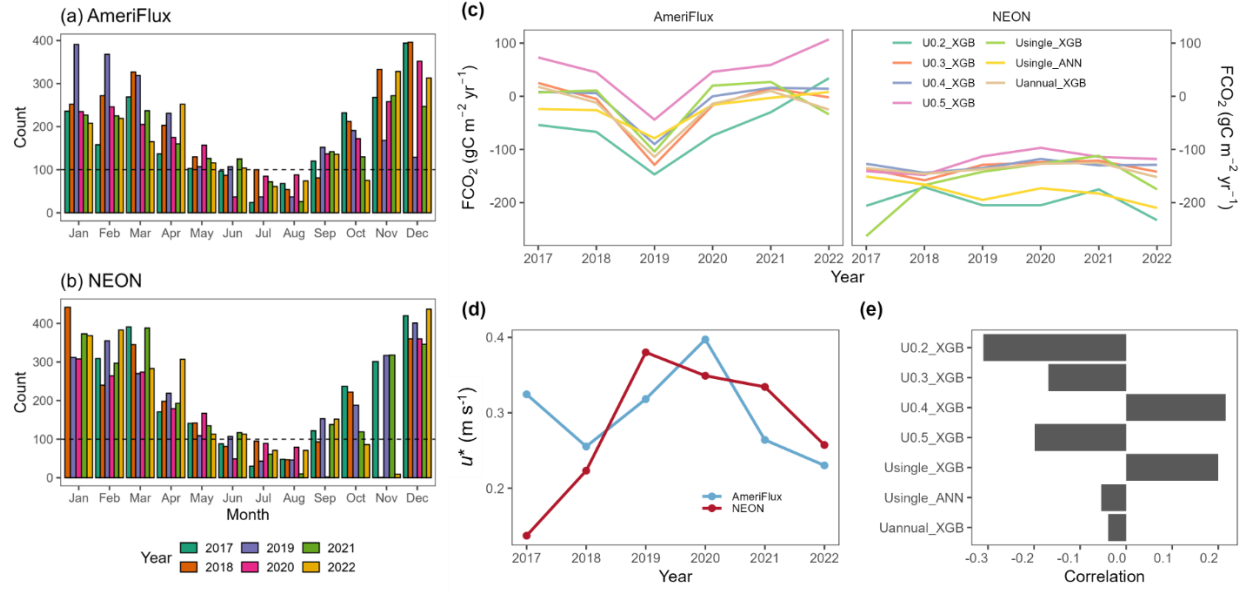


Figure S9. Count of nighttime observations (half hourly) when $u^* > 0.3$ by month for different years for (a) AmeriFlux and (b) NEON; (c) Annual sum of CO₂ flux (FCO₂) calculated at different u^* thresholds. U_{single} refers to a single u^* threshold for the entire measurement period and U_{annual} stands for annual-varying u^* threshold. For U_{single} and U_{annual} , the u^* thresholds are estimated using bootstrapping. We used the 50th percentile (u^*_{50}) of the estimates here. The U_{single} here is u^*_{50} thresholds estimated using bootstrapping, resulting in AmeriFlux ($u^*_{50} = 0.286$ m/s) and NEON ($u^*_{50} = 0.294$ m/s). (d) values for annual-varying u^*_{50} for AmeriFlux and NEON; and (e) correlation of annual sums between AmeriFlux and NEON when using different machine learning algorithms and different u^* threshold.

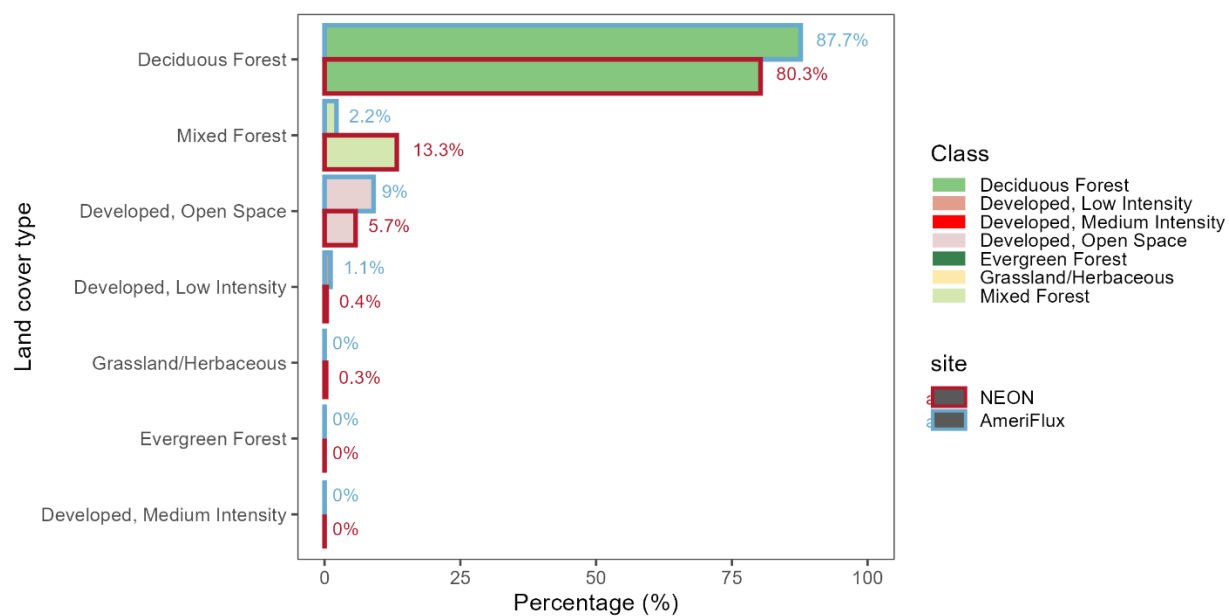


Figure S10. Footprint-weighted percentage of land cover for AmeriFlux and NEON at Bartlett Experiment Forest. AmeriFlux data are shown as bars with blue borders and blue text, while NEON data are shown as bars with red borders and red text.

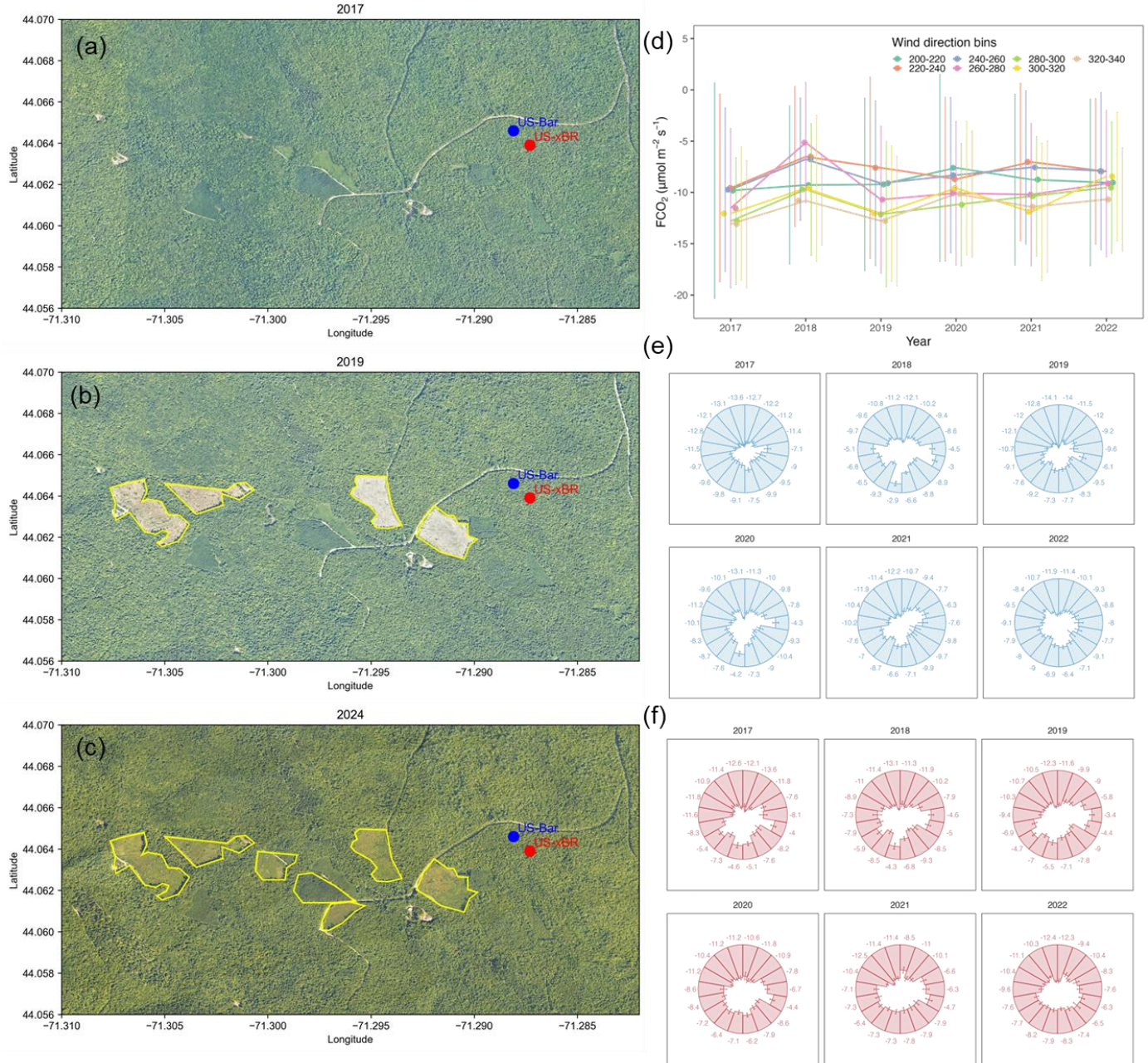


Figure S11. *Top panels*: Management actions at Bartlett Experimental Forest. (a) Harvest in December 2018; (b) Harvest and prescribed fire in May 2021; (c) Image from the spring of 2022. The images were based on NEON high-resolution orthorectified camera imagery mosaic (DP3.30010.001). The location of the eddy covariance towers is marked with a yellow dot. (d) Mean $\pm$ standard deviation of measured FCO₂ during the peak growing season (June, July, and August) by different wind direction bins over time, for the wind direction from the west. *Bottom panels*: Mean $\pm$ standard error of peak growing season measured FCO₂ for different wind directions bins, shown for (e) AmeriFlux and (f) NEON, respectively. Wind direction bins were defined at 20-degree intervals. Reference: NEON (National Ecological Observatory Network). High-resolution orthorectified camera imagery mosaic (DP3.30010.001), provisional data. Dataset accessed from <https://data.neonscience.org/data-products/DP3.30010.001> on November 3, 2025.

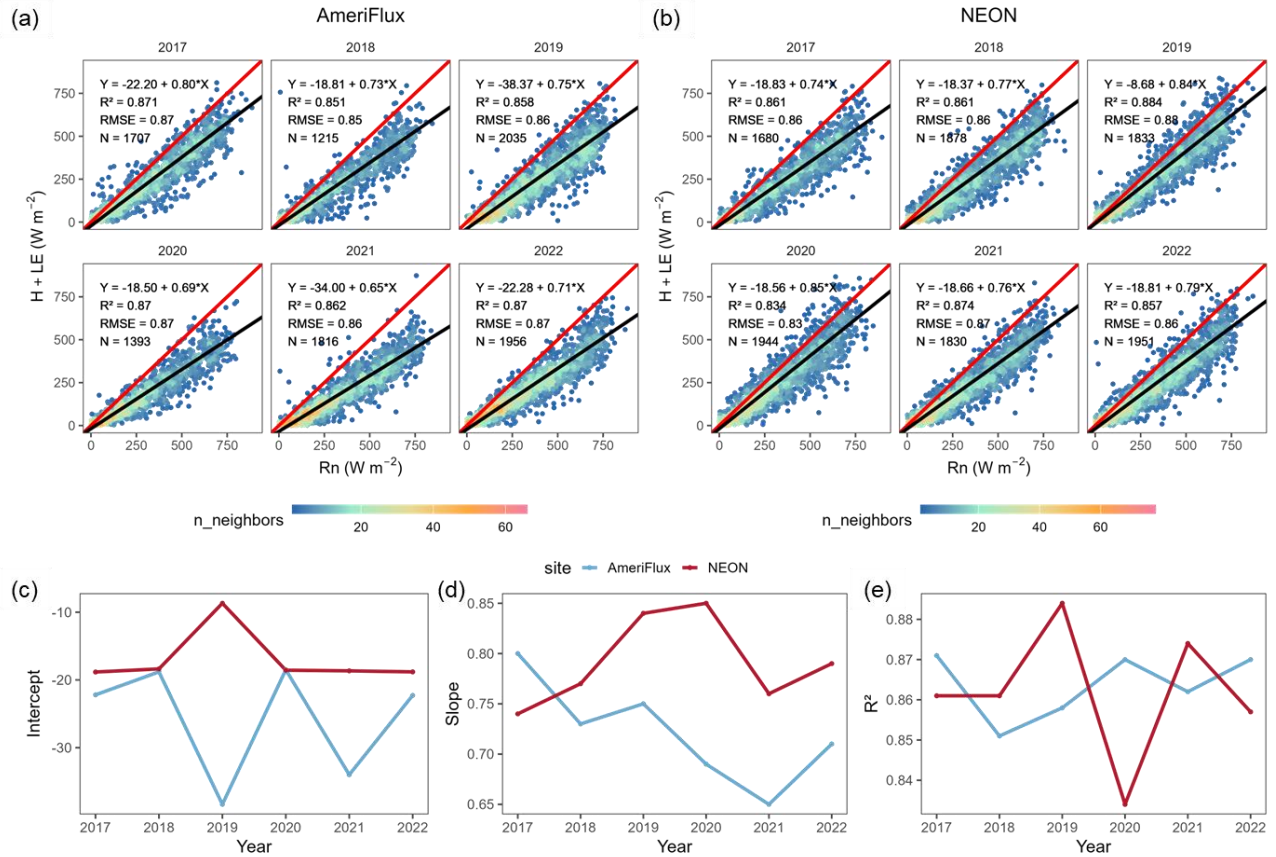


Figure S12. Scatter plots of energy balance closure using daytime data (when incoming PPFD ≥ 5) during the peak growing season months (June, July, and August) across different years: (a) AmeriFlux and (b) NEON. Individual points are colored by the number of neighboring points ($n_neighbors$). (c-e) the intercept, slope, and R^2 between Rn and $H + LE$ evaluated using Model II regression. Rn represents the net radiation. Please note that soil heat flux is not included because it was not measured at the AmeriFlux tower. At the NEON tower, soil heat flux reaches a maximum (early afternoon) of about $+15 \text{ W m}^{-2}$ in May. During June, July, and August, the maximum values are about $+10 \text{ W m}^{-2}$.

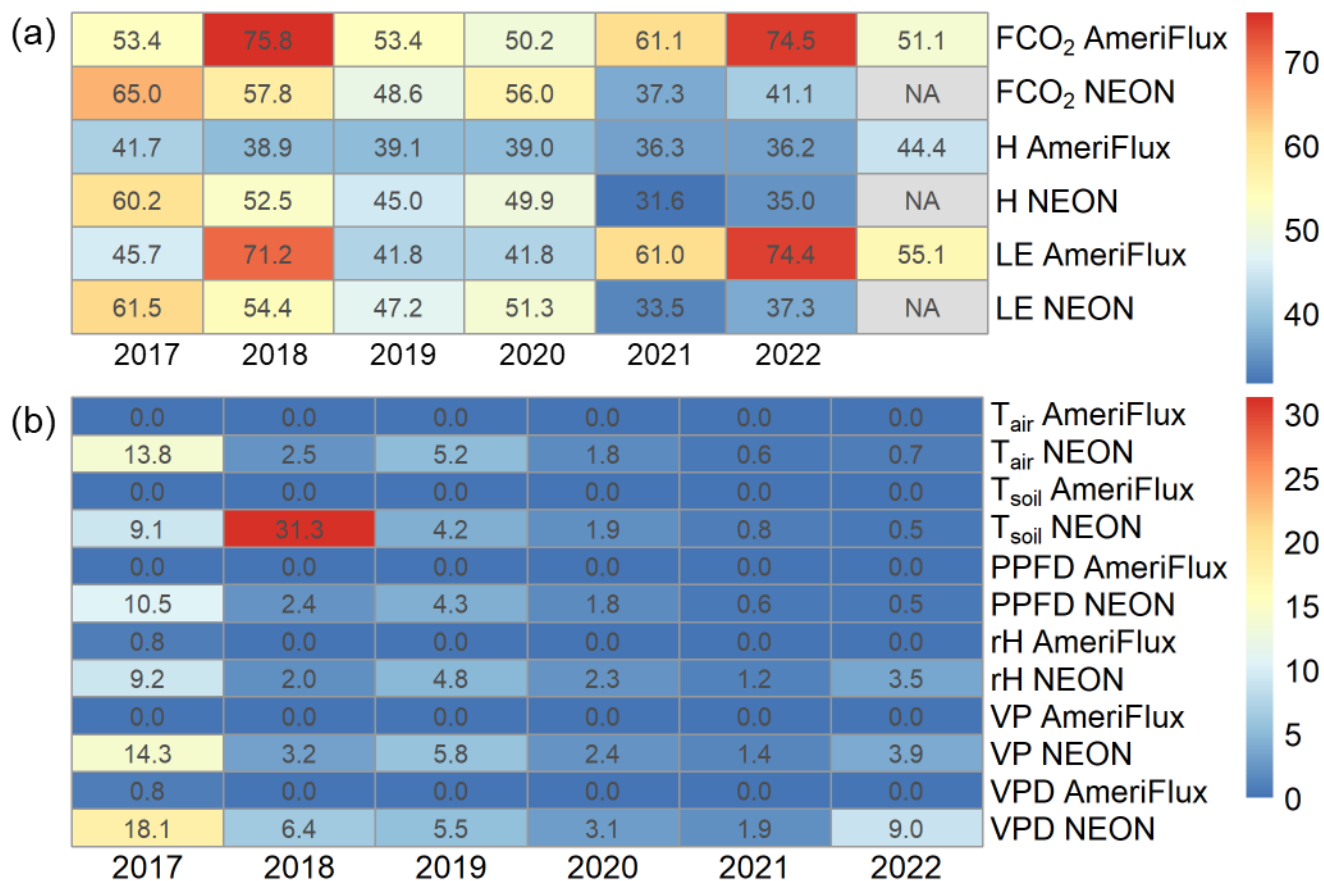


Figure S13. Fraction of missing values (%) in (a) major meteorological variables and (b) flux data for different years from AmeriFlux and NEON measurements, respectively. For flux data, the last column provides information on long-term (2004-2016) data gaps for AmeriFlux measurements.

Abbreviations: air temperature (T_{air}), soil temperature (T_{soil}), photosynthetic photon flux density (PPFD), relative humidity (rH), actual vapor pressure (VP), vapor pressure deficit (VPD), CO₂ flux (FCO₂), sensible heat flux (H), and latent heat flux (LE).

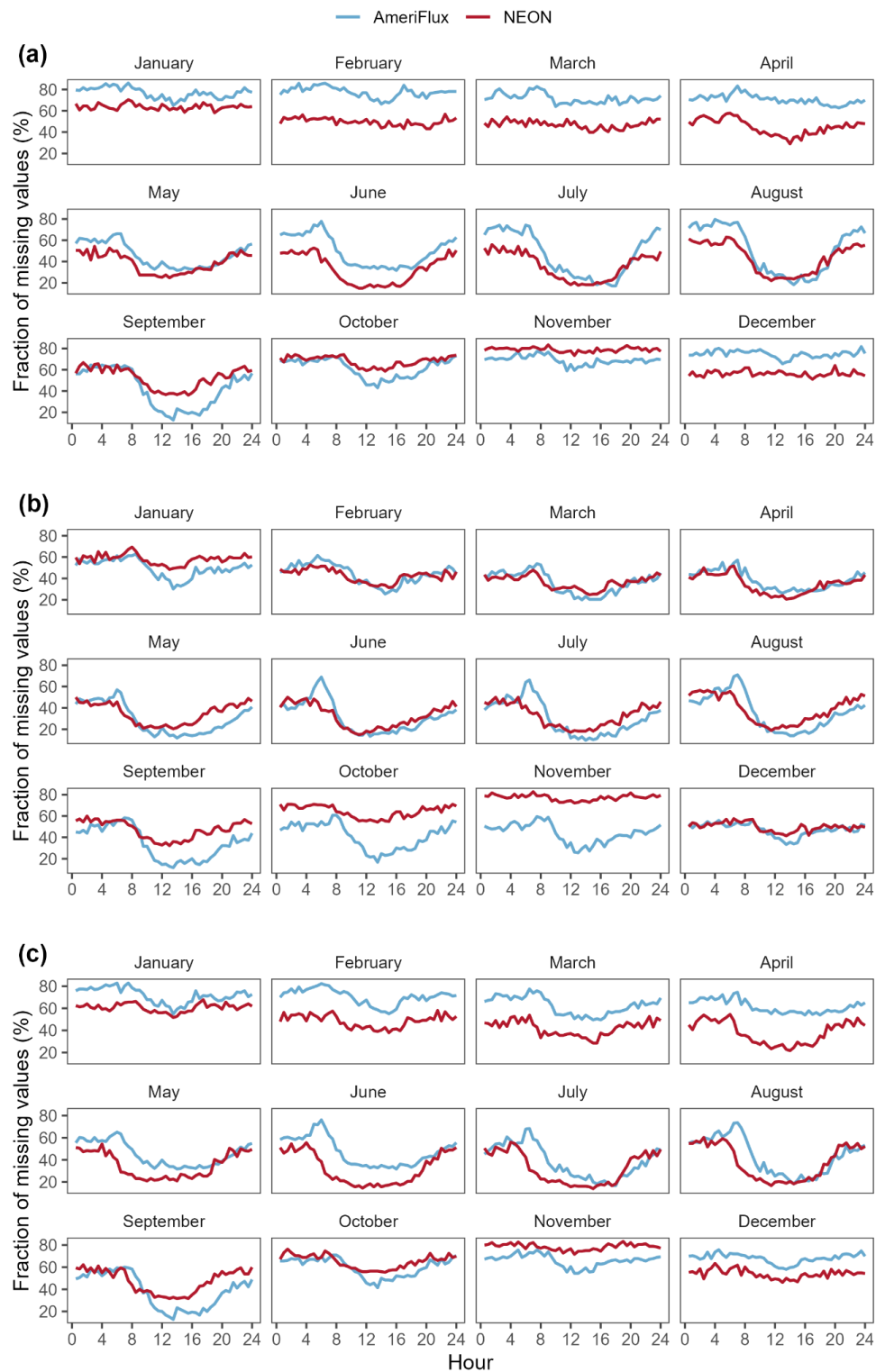


Figure S14. Average diel pattern of fraction of missing values (%) in flux data by month and all years combined for (a) CO₂ flux (FCO₂), (b) sensible heat flux (H), and (c) latent heat flux (LE).

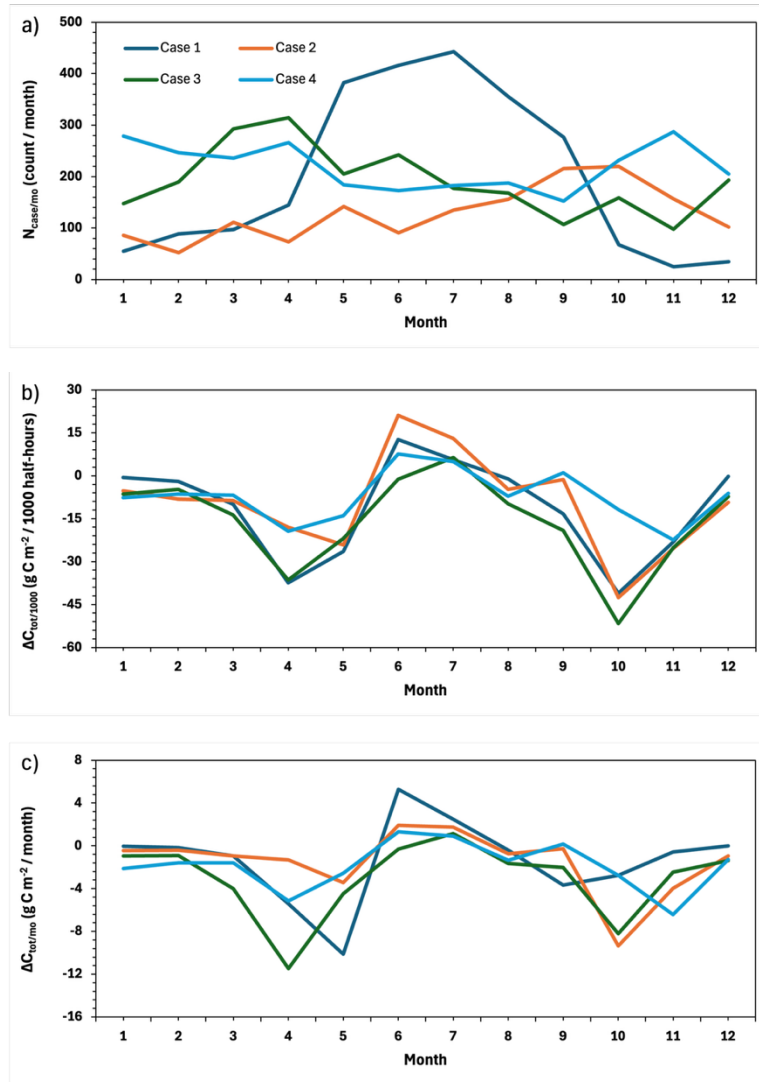


Figure S15. Seasonality of differences in daytime gap-filled and measured values between NEON and AmeriFlux towers at Bartlett Experimental Forest. Cases are as follows: 1) both AmeriFlux and NEON towers had valid measurements; 2) AmeriFlux had a valid measurement, but NEON was filled; 3) NEON had a valid measurement but AmeriFlux was filled; 4) neither tower had a valid measurement and hence both were filled. Panels: a) Monthly breakdown of $N_{\text{case/mo}}$ (count / month), which is the number of half-hourly observations in each case, expressed on monthly basis; b) differences across months in $\Delta C_{\text{tot}/1000}$ (the total difference in C uptake between the NEON and AmeriFlux towers in g C m^{-2} per 1000 half-hourly measurements), by case; and c) the total difference in monthly C uptake ($\Delta C_{\text{tot/mo}}$, $\text{g C m}^{-2} / \text{month}$) for each case ($\Delta C_{\text{tot}/1000} \times N_{\text{case/mo}} / 1000$).