Atmospheric
Chemistry
and Physics

Open Access

Constraining a data-driven CO_2 flux model by ecosystem and atmospheric observations using atmospheric transport

Samuel Upton[1,4]**, Markus Reichstein**[1]**, Wouter Peters**[4,5]**, Santiago Botía**[2]**, Jacob A. Nelson**[1]**,
Sophia Walther**[1]**, Martin Jung**[1]**, Fabian Gans**[1]**, László Haszpra**[6,7]**, and Ana Bastos**[1,3]

[1]Department of Biogeochemical Integration, Max Plank Institute of Biogeochemistry, Jena, Germany
[2]Department of Biogeochemical Signals, Max Plank Institute of Biogeochemistry, Jena, Germany
[3]Institute for Earth System Science and Remote Sensing, Leipzig University, Leipzig, Germany
[4]Environmental Sciences Group, Wageningen University, Wageningen, the Netherlands
[5]University of Groningen, Centre for Isotope Research, Groningen, the Netherlands
[6]International Radiocarbon AMS Competence and Training Center,
HUN-REN Institute for Nuclear Research, Debrecen, Hungary
[7]Institute of Earth Physics and Space Science, Sopron, Hungary

Correspondence: Samuel Upton (supton@bgc-jena.mpg.de)

Abstract. Global estimates of the net ecosystem exchange of CO_2 (NEE) from data-driven models differ widely depending on their underlying data and methodology. Bottom-up models trained on eddy-covariance data are most informative at the ecosystem-level. Top-down models, such as atmospheric inversions, produce regional and global results consistent with the observed atmospheric growth rate, accurately capturing the interannual variability (IAV) of NEE. Both approaches have limitations estimating NEE across scales: Bottom-up models can miss large-scale dynamics of NEE when aggregated globally. Top-down approaches have difficulty relating the large-scale atmospheric signal to biophysical processes at smaller scales. To address these limitations, we create a model that uses a hybrid combination of direct observations and atmospheric dynamics to integrate ecosystem-level eddy-covariance data and atmospheric CO_2 mole fraction data into a single coherent ecosystem-level flux model.

Aggregated globally, our new model estimates an annual sink with a low bias, and consistent IAV when compared with independent estimates. The IAV of the estimated NEE is closer in magnitude to an ensemble of atmospheric inversions, and our model produces a higher temporal coefficient of correlation with these data than state-of-the-art bottom-up data-driven models. This improvement in IAV is achieved without direct access to the observed variability of the atmosphere: the model is trained using only one year of daytime observations from 3 tall-tower observatories. No atmospheric information is available to the model during the production of global NEE estimates. This shows the efficiency of our method in synthesizing top-down information into bottom-up mapping of flux-environment relationships.

1 Introduction

The global net flux of biogenic CO_2 between the land surface and the atmosphere, or net ecosystem exchange (NEE), is a critical but uncertain term in our understanding of the global carbon budget and the climate system (Friedlingstein et al., 2014; Bastos et al., 2022). Current data-driven approaches to modeling NEE can be largely grouped into two categories: top-down approaches that infer NEE from observations of atmospheric CO_2 by prescribing fire and fossil fuel emissions. The second category are bottom-up approaches that model NEE at the ecosystem level from local observations (Kondo et al., 2020). Both approaches provide important information about the function of the biosphere, but have their limitations.

Data-driven bottom-up models are most commonly informed by eddy-covariance observations at the ecosystem scale, along with remotely sensed variables (greenness and water indices, spectral data) (Nelson and Walther et al., 2024; Jung et al., 2020; Kondo et al., 2020; Bodesheim et al., 2018). However, the spatial distribution of NEE and the total magnitude of NEE at global scale remain highly uncertain as the available eddy-covariance data is sparse in important biomes across the globe, such as tropical rain forests (Chu et al., 2017; Hayek et al., 2018; Fu et al., 2018; Jung et al., 2020). Some processes, such as carbon emissions from fires, are not captured by these local observations. Eddy-covariance data are further potentially subject to systematic errors overestimating the carbon uptake (Aubinet et al., 2005), particularly in the tropics, due to complex CO_2 nighttime storage and horizontal advection in the canopy (Fu et al., 2018; Moncrieff et al., 1996). Historically, when these eddy-covariance data are used to train a data-driven bottom-up flux model for global upscaling of NEE (e.g. FLUXCOM V1, Jung et al., 2020), the underlying issues of data availability and representation are propagated into the global result resulting in an overestimation of the tropical carbon sink and by extension, the global carbon sink. Therefore data-driven global products such as FLUXCOM V1 NEE, which are very dependent on the quality and completeness of the training data, are not fully consistent with the magnitude of the growth rate of atmospheric CO_2, or the interannual variability (IAV) of the global NEE signal (Jung et al., 2020; Kondo et al., 2015). However the length of the available eddy-covariance record has increased, and the data collection and processing have improved (Pastorello et al., 2020). These longer, improved data, along with improved handling and gap-filling of the complementary co-located remotely sensed variables (Walther and Besnard et al., 2022) have allowed for FLUXCOM X-BASE to improve its estimates of the magnitude of regional and global NEE. Despite these improvements, the eddy-covariance record may miss complex drivers of IAV, or they may be obscured by sensor issues (Jung et al., 2024). The under-estimation of the global NEE IAV in FLUXCOM X-BASE remains unresolved (Nelson and Walther et al., 2024).

Top-down approaches, most commonly atmospheric inversions, are trained using observations of CO_2 mole fraction from tall-tower observatories and/or satellite retrievals of CO_2. Atmospheric inversions use a Bayesian inversion framework and an atmospheric transport model to produce global or regional estimates of NEE which are consistent with the observed atmospheric signal (Chevallier et al., 2005; Peylin et al., 2013; Crisp et al., 2022; Ciais et al., 2022). These inversion systems by design provide estimates of NEE that agree with the growth rate of CO_2 at the global scale and across broad latitudinal bands (Peylin et al., 2013) despite differences in priors and representation of transport mechanisms. Atmospheric inversions are formulated to capture the structure and dynamics of atmospheric CO_2, and can reproduce the observed IAV of NEE with very high accuracy (Rödenbeck et al., 2018).

However, despite improvements in the ability of regional inversions to provide spatially explicit flux estimates (Munassar et al., 2022), global top-down estimates of NEE lack the ability to spatially map the atmospheric signal to local biophysical conditions. Global inversions are more commonly used to understand the land surface at larger integrated scales, and are not directly comparable with eddy-covariance data (Kaminski and Heimann, 2001). Like bottom-up systems, the top-down approach is also limited by a lack of observations. The observational network does not provide sufficient tropical coverage for a robust top-down estimate of the tropical land flux (Palmer et al., 2019). In the Southern Hemisphere, the weaker gradients in CO_2, and shorter observational records also exacerbate issues of data availability and data quality (Peylin et al., 2013). Additionally, in tropical regions there are less independent data to validate atmospheric inversion estimates (Chevallier et al., 2019).

Previous work has demonstrated that top-down information can be effectively combined with a bottom-up data-driven flux model, improving the regional and global performance by creating a dual constraint from eddy-covariance data and atmospheric inversion estimates of NEE (Upton et al. 2024). Although this dual-constrained model was able to infer regional and global NEE integrals with a much lower bias compared with other best estimates of NEE than the comparable FLUXCOM RS+METEO V1 results (Jung et al., 2020), the model has several important limitations; First, the atmospheric constraints were pre-calculated using large-scale mean NEE from an ensemble of inversions, rather than from atmospheric CO_2 itself. Second, the atmospheric constraint was connected to the local bottom-up data-driven flux model using a static, statistical model rather than an atmospheric transport model. This means that the additional atmospheric information is aggregated, and adds no additional information on the spatial distribution of NEE. Both issues limited the amount of information available to the model from the atmospheric constraint.

To address these limitations, we create a new model that uses a hybrid combination of direct observations and

atmospheric dynamics to integrate ecosystem-level eddy-covariance data and atmospheric CO$_2$ mole fraction data into a single coherent ecosystem-level flux model. To create the computational link between ecosystem-level fluxes and atmospheric CO$_2$ mole fraction we use a physical atmospheric transport model, the Stochastic Time-Inverted Lagrangian Transport (STILT) model (Lin et al., 2003). In this way, both eddy-covariance and atmospheric observations are used to train and update the model at the ecosystem level, the result is a bottom-up model that uses atmospheric data to improve the representation of the land surface. This technique opens a large corpus of new observations for constraining the model. Due to varying transport patterns around each tower, the data-driven model is exposed to a large number of non-eddy-covariance locations which represent biophysical conditions that are coupled to important dynamics in global NEE. This enrichment of the model's view of the underlying relation between driver variables and NEE, we hypothesize, allows the system to learn additional dynamics in global, and regional NEE.

2 Data

2.1 Ecosystem-level data

At the ecosystem level, we define Net Ecosystem Exchange (NEE) as the simple difference between Gross Primary Production (GPP) and Total Ecosystem Respiration (TER). These terms are measured as the flux density $\mu mol\,CO_2\,m^{-2}\,s^{-1}$. The sign of NEE indicates the direction of the flux. When NEE is negative, GPP exceeds TER and the flux is a local sink of CO$_2$. If NEE is positive, TER exceeds GPP and the ecosystem is a local source. NEE represents the net carbon exchange of the vegetation and does not include disturbance fluxes, fires, or out-gassing from freshwater ecosystems.

The core ecosystem-level meteorological and NEE data was collected between 2001 and 2020 at 294 globally distributed eddy-covariance (EC) towers (see appendix for a full list of sites). These data were then processed using the ONE-FLUX processing pipeline (Pastorello et al., 2020). Variables from the eddy-covariance sites are air temperature, vapor pressure deficit, incoming shortwave radiation, the computed potential shortwave incoming radiation, and the computed wind speed.

During training, at pixels co-located with eddy-covariance towers we use the surface reflectances (MCD43A4, Schaaf and Wang, 2015a) and land surface temperature (LST) from the MOderate Resolution Imaging Spectroradiometer (MODIS) collection v006 (http://daac.ornl.gov/MODIS/, last access: 11 January 2024). The surface reflectances are used to compute two vegetation indices, the enhanced vegetation index (EVI) (Huete et al., 2002), and the spectral reflectance of vegetation in the near-infrared (NIRv) (Badgley et al., 2019). A water index, the normalized difference water in-

dex (NDWI) with MODIS band 7 (Gao, 1996) was also included. MODIS data has a spatial resolution of 0.05°, and a 1 d temporal resolution. The full set of variables is described in Table 1.

When producing our atmospheric constraint, and when estimating global NEE, the drivers for our bottom-up model are the same set of remotely-sensed and eddy-covariance quantities, extracted from remotely-sensed and meteorological reanalysis data. The RS data are derived from the 0.05° global MODIS product (MCD43C4, Schaaf and Wang, 2015b), and the global meteorological variables are derived from the European Center for Medium-Range Weather Forecasting (ERA5) atmospheric reanalysis data (https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5, last access: 11 January 2024). ERA5 is provided at 0.25° spatial resolution, and hourly temporal resolution (Hersbach et al., 2020). Air temperature at 2 m height, incoming shortwave radiation, and vapor pressure deficit (computed from air temperature, relative humidity, and surface pressure) are used to create the global meteorological data.

All data are accessed through the FLUXCOM-X code base using the preprocessing, and gap-filling from Nelson and Walther et al. (2024) and the procedures of FluxnetEO data version 2 (Walther and Besnard et al., 2022), and the quality flagging described in Jung et al. (2024). Please see Nelson and Walther et al. (2024) for a full description of the FLUXCOM-X data and processing environment.

2.2 Atmospheric data

The model in this study uses observations of atmospheric CO$_2$ mole fraction from tall-tower observatories measured at three sites: ATTO (Botia et al., 2022), Hegyhatsal (Haszpra, 2024) and Zottino (Tran et al., 2024a). These sites are chosen to include one tropical, one extra-tropical and one boreal domain. The mole fraction data used for the Amazon Tall Tower Observatory in Brazil (ATTO), collected at 79 m are the same used and fully described in Botia et al. (2022). Mole fraction data for the Hegyhátsál tower in Hungary (HUN), collected at 115 m are provided by the Integrated Carbon Observation System (ICOS). The observations for the Zotino Tall Tower Observatory in Siberia (ZOTTO), collected at 301 m are provided by Tran et al. (2024b). The data for all three towers was provided hourly, for one calendar year (2019 for ZOTTO and Hegyhátsál, 2015 for ATTO). Only daytime observations from 13:00–17:00 (local time) were used. The selected times represent a daytime planetary boundary layer (PBL) with assumed well-mixed convective conditions (Peters et al., 2010).

2.3 Reference data

For global and regional comparison we use the land flux corrected for fossil fuel and cement production from the ensemble of $N = 14$ atmospheric inversions from GCB23

Table 1. Driver variables for the model (EC-STILT) in this study.

Variable	Short Name	Unit	Source
Air Temperature	TA	°C	EC/ERA5
Vapor Pressure Deficit	VPD	hPa	EC/ERA5
Incoming Short Wave Radiation	SW_IN	$W\,m^{-2}$	EC/ERA5
Potential Incoming Short Wave Radiation	SW_IN_POT	$W\,m^{-2}$	ERA5
Daytime Land Surface Temperature	LST	°C	MODIS
Daytime Land Surface Temperature	LST	°C	MODIS
Near-infrared Reflectance of Vegetation	NIRv	–	MODIS
Enhanced Vegetation Index	EVI	–	MODIS
Normalized-difference Water Index	NDWI	–	MODIS

(https://doi.org/10.18160/4M52-VCRU, Luijkx et al., 2024, last access: 4 March 2024). The atmospheric inversions are provided as monthly data with a 1° spatial resolution. See Friedlingstein et al. (2023) for a full description of the ensemble members. These fluxes are adjusted for fire by removing the fire emissions from Global Fire Emissions Database, Version 4.1 (GFEDv4) (last access: 13 February 2024) (Randerson et al., 2017).

For comparison with the current state-of-the-art bottom-up NEE model, we use the global NEE product from the latest version of FLUXCOM X-BASE (Nelson and Walther et al., 2024, last access: January 2025). The FLUXCOM X-BASE NEE model (hereafter X-BASE) is trained using the same data processing and data quality flagging, including gap-filling. X-BASE uses a different data-driven algorithm, XGBoost (Chen and Guestrin, 2016), instead of a neural network. Comparison with X-BASE allows us to understand the performance of our system relative to a highly optimized bottom-up model. So, despite differences in structure, we consider X-BASE to be the most appropriate bottom-up model for comparison.

For regional analysis we use the set of 11 large regions from the TransCom 3 intercomparison project (TransCom) project (Baker et al., 2006). See Fig. A1 for the coverage and names of the 11 land regions used in this study. For analysis of model function by climate zone, we use the Köppen-Geiger (KG) classification system using maps from Beck et al. (2018), for the years 1991–2020. The Köppen-Geiger classification groups the land-surface into regions by their annual climate (Fig. A2).

3 Methods

In this work, we start with a neural network model which uses eddy-covariance, meteorological, and remotely sensed data to estimate NEE at the ecosystem level. We create a hybrid objective function which allows direct constraint by eddy-covariance (EC) observations and hybrid constraint from direct mixing ratio observations of CO₂ using the Stochastic Time-Inverted Lagrangian Transport (STILT) at-

mospheric transport model as the computational link from the EC-model to the top-down observations. We therefore refer to this new data-driven system with its dual constraints as EC-STILT. Its components and their recombination in a machine-learning framework are described below.

3.1 Ecosystem-level model

The ecosystem-level model takes as input observations of meteorological drivers (from eddy-covariance or reanalysis data) and remotely-sensed drivers and predicts NEE in $\mu mol\,CO_2\,m^{-2}\,s^{-1}$ (Fig. 1, red lines, orange lines) at an hourly tempo. As a machine-learning system, the ecosystem-level model can be described as a feed-forward neural network trained using standard backpropagation techniques (Kelley, 1960). The network is a set of fully-connected layers which consist of nodes or "neurons". The neurons are exposed to the output of all neurons in the previous layer. Between each layer of neurons, the output is passed through a non-linear activation function. Additionally, the output of the activation function is passed through a normalization function, which scales the output to a mean of 0 and a standard deviation of 1. Our network is a set of three fully-connected layers with the Rectified Linear Unit (ReLU) activation function (Agarap, 2019) and batch normalization. Each layer in a neural network is a complex, non-linear embedding of the previous inputs into a new latent space. This series of embedding functions is the learning space of the network. These latent spaces allow the model to learn complex non-linear relationships between driver variables and NEE. Because these relationships are discovered in these abstract non-geographic spaces, they can represent complex space-for-time and time-for-space substitutions. We refer to the process of learning these non-linear, non-geographic mappings between the input distribution and NEE as learning in "environmental response" space.

3.2 STILT transport

The computational link from this ecosystem-level model to the atmospheric constraint (Fig. 1, purple lines) is built
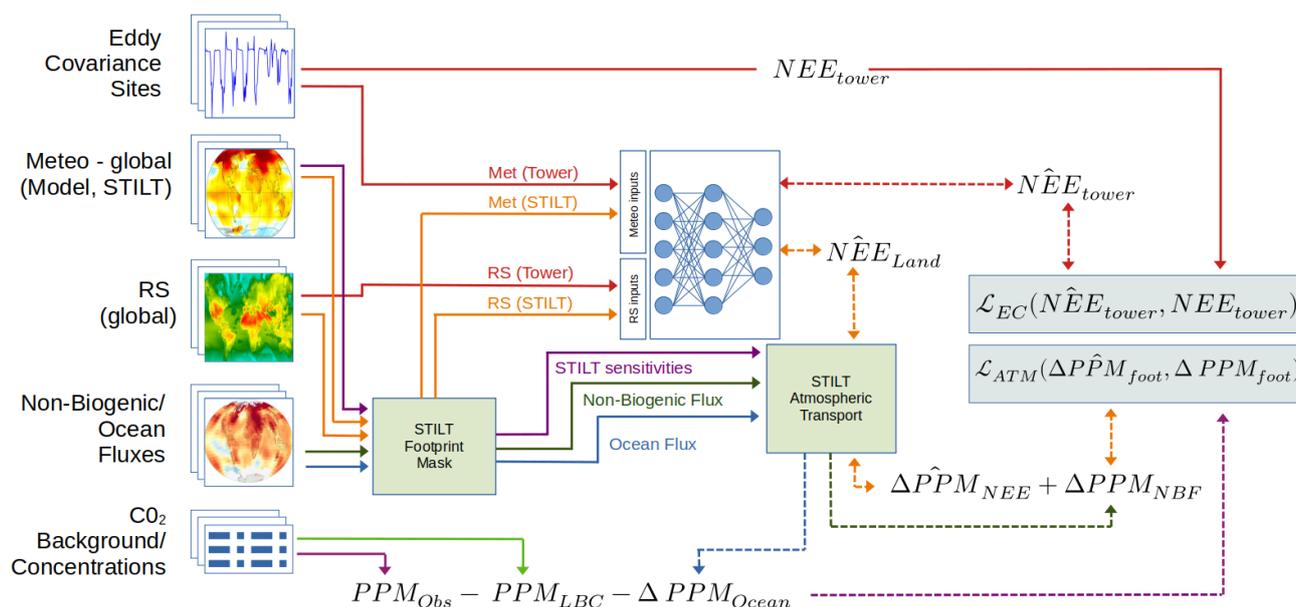
Figure 1. Data flow for atmospheric constraint calculation. An objective term from eddy-covariance (red lines) is created from tower observations of driver variables, co-located remotely sensed data, and tower observations of NEE. An objective term from atmospheric observations (model: orange lines, ancillary data: other colors) uses the STILT model to transport modeled flux densities into concentration space, and corrected for non-local and non-biogenic CO_2: The lateral boundary condition (PPM_{LBC}), or background CO_2 from the well-mixed atmosphere, and any ocean flux contributions (ΔPPM_{Ocean}) are removed from the tower observation (PPM_{Obs}) to create the observed value for the contribution from the footprint (ΔPPM_{foot}). Non-biogenic fluxes (ΔPPM_{NBF}) from fire and fossil fuels are added to the predicted NEE from the footprint ($\Delta\widehat{PPM}_{NEE}$) to produce the predicted contribution from the footprint ($\Delta\widehat{PPM}_{foot}$). Dashed lines indicate terms that are created during the process of training. See Sect. 3.3 and 3.4 for a full description of terms. Lines with bi-directional arrows indicate the flow of backpropagation leading to the weights of the neural network being updated.

around the STILT atmospheric transport model (Lin et al., 2003). STILT is a Lagrangian particle dispersion model (LPDM) which uses meteorological forcing data to compute the transport of sources and sinks to the time and location of an observation. The STILT footprint represents this atmospheric transport as a set of sensitivities of the final observation to fluxes at different times and locations. This footprint can be multiplied with a set of local estimates of NEE, which transports them and converts them into a simulated mole fraction. EC-STILT uses a set of pre-computed STILT footprints at the three tall tower locations, for midday observations, for one year (Fig. 1). For each tower, the STILT model is run for each selected observation over the year, producing footprints covering the tower domain hourly for the 10 d prior to the observation time. STILT domains (Fig. 2) are selected to capture the relevant near field over the time of analysis based on regional meteorological conditions (Lin et al., 2003).

For the atmospheric constraint, we must isolate the biogenic contribution (ΔPPM_{NEE}) of an hourly observation of CO_2 (PPM_{Obs}). Conceptually, the biogenic contribution is the residual of the observation and the non-biogenic terms: the lateral boundary condition (LBC), non-biogenic fluxes (NBF) and Ocean contributions. This is presented in Eq. (1),

with spatially explicit terms on the right.

$$PPM_{Obs} - PPM_{LBC} - \Delta PPM_{Ocean} = \Delta PPM_{NEE} + \Delta PPM_{NBF} \quad (1)$$

The NBF contribution (ΔPPM_{NBF}) consists of fossil fuel fluxes and fire emissions. For fossil fuel emissions, we use fossil fuel/energy signal across sectors, using GridFED (version 2022.2, last access: 15 February 2024) (Jones et al., 2021), provided at a 0.1° resolution. Fire emissions are provided by the Global Fire Emissions Database, Version 4.1 (GFEDv4) (last access: 13 February 2024) (Randerson et al., 2017) provided at a 0.25°. For the GFEDv4 data, temporal profiles were used to add monthly, daily, and diurnal cycles to the fire signal. The GridFED data are interpolated to the 0.25° spatial resolution of the analysis. We do not include lateral transport fluxes in the current analysis. Crop and wood harvest are important for regional and long term accounting (Ciais et al., 2022), but are not critical for the instantaneous carbon budget that is represented in Eq. (1). The potential impact of riverine transport are discussed below in Sect. 5.1.

The lateral boundary condition of the region (PPM_{LBC}) is precomputed using The Jena CarboScope (s04ocv4.3) (Rödenbeck et al., 2003), following Botia et al. (2022). Optimized results from the CarbonTracker Data Assimilation Shell (CTDAS) (van der Laan-Luijkx et al., 2017) (CTE2020) are used for the ocean flux and its conversion
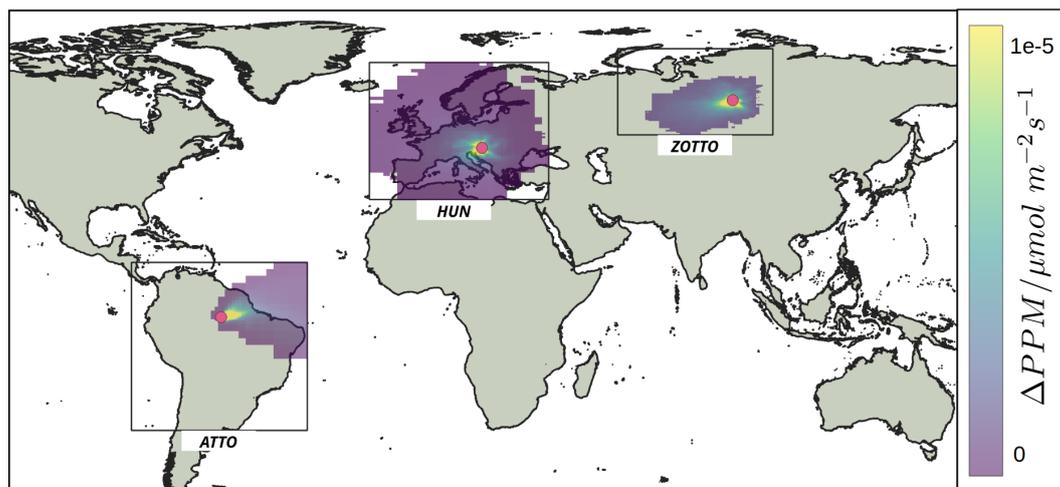
Figure 2. Locations of tall towers used by the atmospheric constraint (red dots). The boxes represent the near-field domain where STILT is computed, and the grid within each domain is the mean annual sensitivities of the computed STILT footprints. The colorbar, in ΔPPM $(\mu\mathrm{mol}\,\mathrm{m}^{-2}\,\mathrm{s}^{-1})^{-1}$, represents the mean sensitivity of the tower observation to the region over the full year.

to mole fractions (ΔPPM$_\mathrm{Ocean}$) with STILT. The LBC values are calculated for each tall tower for each observation time. We used the ensemble of STILT trajectories released at each data point to obtain the mean ending position (lat, lon) as well as a mean ending height above the ground. Using this information we sample global 3D fields, in this case the optimized CO$_2$ mole fractions from CarboScope, and obtain a mole fraction associated with a measurement at the tower. We acknowledge that this method is sensible to biases in the global 3D fields, but for example at ATTO a bias-corrected version of the LBC yielded very similar results to the Carbo-Scope LBC, please see Botia et al. (2022) for the full discussion.

3.3 Objective function – eddy covariance

During training, EC-STILT generates an objective function with two terms, or losses: the ecosystem loss term $\mathcal{L}_\mathrm{EC}$, and the atmospheric loss term $\mathcal{L}_\mathrm{ATM}$. To generate the ecosystem loss term $\mathcal{L}_\mathrm{EC}$ from eddy-covariance, the EC-model $\mathcal{F}()$ is run using a batch of data randomly selected by time and site, x_batch collected at eddy-covariance tower sites (Eq. 2, along with remotely sensed (RS) data at pixels that are co-located with the eddy-covariance tower, producing an estimate of NEE at the ecosystem level (Eq. 2). The ecosystem loss term is computed as the mean squared error (MSE) between this estimate and the observed NEE from the eddy-covariance observations (Eq. 3).

$$\widehat{\mathrm{NEE}} = \mathcal{F}(x_\mathrm{batch}) \tag{2}$$

$$\mathcal{L}_\mathrm{EC} = \mathrm{MSE}(\mathrm{NEE}_\mathrm{obs}, \widehat{\mathrm{NEE}}_\mathrm{tower}) \tag{3}$$

3.4 Objective function – atmospheric

To generate the atmospheric loss term $\mathcal{L}_\mathrm{ATM}$ for each training step, an hourly daytime (13:00 to 17:00 local time) observation of the mole fraction of CO$_2$ for each tower is selected. For each observation, a pre-computed STILT footprint, along with associated NBF, LBC and ocean data are retrieved. The observed mole fraction, LBC and ocean contributions are used to create the left-hand side of Eq. (1). The ocean term represents the flux from any pixels under the footprint, transported using the footprint into concentration enhancements at the tower. To create the right-hand side of Eq. (1), which is the change in CO$_2$ mole fractions attributable to fluxes within the footprint (ΔPPM$_\mathrm{foot}$). The ecosystem-level model is then run for each non-zero location in the STILT footprint, producing an estimate of local NEE (Fig. 1, NEE$_\mathrm{land}$). The NEE inferences and NBF values are transported with the footprint into concentration enhancements at the tower. This produces the two terms ΔPPM$_\mathrm{NEE}$, and ΔPPM$_\mathrm{NBF}$ which are then added element-wise. These sums represent the total simulated change in PPM at the tower. The loss term $\mathcal{L}_\mathrm{tower, ATM}$ is the squared difference between the simulated change in concentration $\widehat{\Delta\mathrm{PPM}}_\mathrm{foot}$, and the observed change in concentration ΔPPM$_\mathrm{foot}$. For a single training step, a loss is calculated for one observation for every tower. These three tower losses are averaged, producing the final atmospheric loss term (Eq. 5).

$$\mathcal{L}_\mathrm{tower, ATM} = (\Delta\mathrm{PPM}_\mathrm{foot} - \widehat{\Delta\mathrm{PPM}}_\mathrm{foot})^2 \tag{4}$$

$$\mathcal{L}_\mathrm{ATM} = \frac{\sum_t^{\mathrm{Num.\,Towers}} \mathcal{L}_\mathrm{t, ATM}}{\mathrm{Num.\,Towers}} \tag{5}$$

The two terms of the objective function, $\mathcal{L}_\mathrm{EC}$ and $\mathcal{L}_\mathrm{ATM}$ are combined using a learned weighting scheme (Kendall et al., 2018). This allows the model to learn the appropriate rela-

tive weight to assign each loss term during training, based on their internally estimated uncertainty. To achieve this, the model has a set of additional parameters which are used to learn an estimate of the homoscedastic uncertainty for each term of the objective function. This uncertainty is dependent on the inherent noise in the training data, rather than the quality or quantity of training data. For EC-STILT these mixing parameters, σ^2_{EC} and σ^2_{ATM}, are added to the processing chain of the model after its initialization. During training, using the normal backpropagation process that uses the chain rule to attribute and update the free parameters of the neural network according to their contribution to the loss value, the model also updates these mixing parameters. For the individual tasks $L \in$ [EC, ATM], the σ^2_L parameter is used to create two terms; w_L (Eq. 6), and s_L (Eq. 7). These are then used to calculate the effective loss, balanced by the learned uncertainty of the terms (Eq. 8). The evolution of these terms during training is provided in Fig. A4. After this final weighting calculation, the loss is combined in small batches to stabilize model training during backpropagation.

$$w_{EC} = \frac{1}{2\sigma^2_{EC}}, \ w_{ATM} = \frac{1}{2\sigma^2_{ATM}} \tag{6}$$

$$s_{EC} = \log\sqrt{\sigma^2_{EC}}, \ s_{ATM} = \log\sqrt{\sigma^2_{ATM}} \tag{7}$$

$$\mathcal{L}_{total} = (w_{EC} \times \mathcal{L}_{EC}) + (w_{ATM} \times \mathcal{L}_{ATM}) + s_{EC} + s_{ATM} \tag{8}$$

3.5 Model training

EC-STILT is trained using 10-fold cross-validation. The full set of eddy-covariance observations are split by tower location into 10 equal subsets. Ten ensemble members are created, each trained on 8 of the subsets, with one held out for validation and one held out for testing. The ensemble is used to define the flux uncertainty as given in the Results section, as a 1σ of the member spread. The available footprints for each tower were randomized and 80 % are used for training, 20 % for validation on each member. The cross-validation results for the eddy-covariance level results below are the ensemble member results against the 10 % of sites held out for testing.

The STILT footprints were run for a 0.25° grid. All MODIS driver variables and NBF data were translated to this grid. MODIS variables were averaged. Where necessary, NBF fluxes were interpolated using the nearest neighbor method which reduced the overestimation of emissions in pixels surrounding urban hotspots. When run globally, EC-STILT is run at full MODIS 0.05° resolution.

3.6 Model evaluation

We perform several analyses by KG geophysical region to understand how model performance is modified by the atmospheric information available during training. To discover the regions which contribute to the global IAV in EC-STILT,

we use the covariance method described in Lee et al. (2023). This method uses the row sum of the covariance matrix, scaled by the sum of the full covariance matrix to estimate the per-pixel contribution to the IAV, which is in turn additive. We then sum by KG region to determine the relative contribution.

To test our hypothesis that adding an atmospheric constraint improves the representation of the land surface, we estimate the probability distribution functions (PDF) of the model's training data (such as temperature and VPD) by KG region. We also estimate the PDF of the full land surface, as represented by the full dataset used to create a global multiyear estimate of NEE. For the training data we estimate two different PDFs: one PDF of the data provided from the eddy-covariance towers, which is the training set for X-BASE, and one PDF representing additionally the areas of the land-surface which are under the STILT footprints, which is the training set of our model. The footprint data is weighted by the number of times that time and location is used during training. Global and regional PDFs are generated from a random subsample of the full dataset. The subsamples are the per-year time series for ten percent of randomly selected spatial locations. To quantify the relationship between the training PDFs and the full PDF, we use two metrics that describe the distance between two distributions; Kullback-Leibler and Jennsson-Shannon.

To understand the impact of the STILT data on our results, we try to isolate and test specific potential confounders by training nearly identical models with specific modifications. These models share the same architecture and random state. To test the role of model architecture on model performance, we train an EC-only version of our model which lacks the STILT operator, and so has no additional atmospheric information during training. To test for potential influences from the atmospheric inversion data which we use to calculate the LBC timeseries for each tall tower, we train a model which replaces the hourly LBC with the annual mean LBC for each time step. In this way this mean-LBC model has no time-varying information on the background state of the atmosphere.

4 Results

4.1 Global model performance

EC-STILT produces an estimate of annual global NEE of $-3.90 \pm 3.63\,\mathrm{Pg\,C\,yr^{-1}}$, compared to the ensemble mean of atmospheric inversions in GCB23, adjusted for fire, $-4.29 \pm 0.65\,\mathrm{Pg\,C\,yr^{-1}}$ (Fig. 3B). The RMSE of the EC-STILT annual mean with the inversion ensemble annual mean is $0.69\,\mathrm{Pg\,C\,yr^{-1}}$, compared with $1.81\,\mathrm{Pg\,C\,yr^{-1}}$ for X-BASE. The uncertainty of the global annual flux is very high (Fig. 3A), being driven by uncertainties in tropical NEE (Fig. 4C).
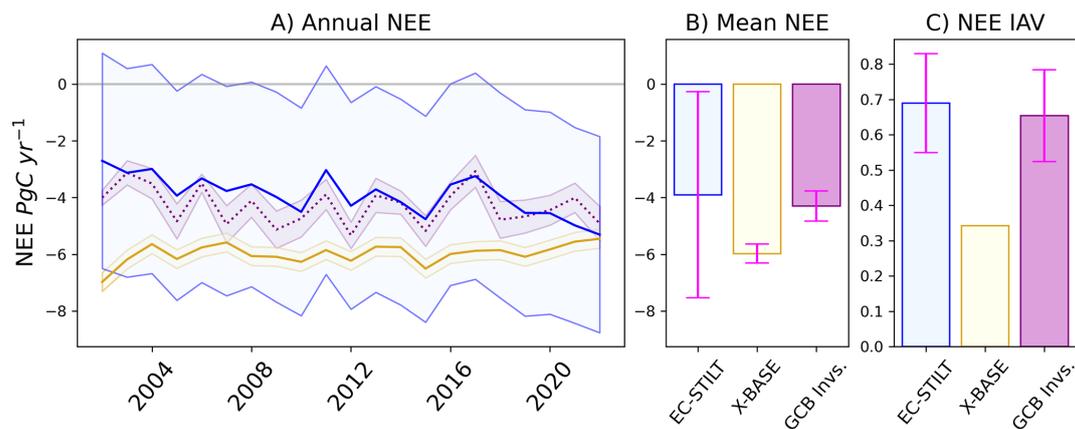
Figure 3. (A) Annual NEE integrated globally in $PgC\,yr^{-1}$. The blue line is the mean of the EC-STILT ensemble with the shaded area representing the 1σ range across ensemble members. The yellow line is FLUXCOM X-BASE. The purple dotted line is the mean of the GCB23 ensemble of atmospheric inversions, and the shaded area is the standard deviation across the ensemble members. **(B)** Mean NEE integrated globally over all years. The error bars represent the 1σ range across ensemble members or published uncertainty. **(C)** IAV of NEE, calculated as the standard deviation of the global integral. The error bars represent the 1σ range of IAV across ensemble members.

EC-STILT has learned a relationship between the input variables and NEE, which captures a large part of the natural interannual variability of global annual NEE. When producing an estimate of global NEE, the model takes the driver variables across the full land surface as inputs, but does not access any STILT footprint data, LBC or NBF data, or mole-fraction data, which are only used in constraint during training. When the model is run globally for years 2001–2021, the standard deviation of annual NEE (IAV) of the EC-STILT member mean, $0.69\,PgC\,yr^{-1}$ (1σ), is substantially closer to the GCB23 inversion ensemble IAV of $0.65\,PgC\,yr^{-1}$ than that of X-BASE (Fig. 3C). The R^2 of EC-STILT annual mean NEE with the GCB23 inversion ensemble is 0.42 ($N = 21$ years), compared with an R^2 of 0.02 for X-BASE. EC-STILT IAV is consistent across the model members despite the large uncertainty in the magnitude of the flux. The mean R^2 between the 10 members and GCB23 inversion ensemble is 0.34 (1σ 0.08), and the mean IAV is 0.71 (1σ 0.14) $PgC\,yr^{-1}$.

Spatially, the distribution of the mean annual NEE is largely consistent with X-BASE (Fig. 4B) with a spatial correlation of 0.71 for global mean NEE. EC-STILT has a stronger Amazonian sink and weaker boreal sink (Fig. 4C). The boreal reduction is a reduction in the length and intensity of the growing season. The EC-STILT results have removed several hotspots of source, in the Sahel and in the Indian subcontinent. These potentially unrealistic hotspots in X-BASE can be attributed to an incorrect learned relationship for crop-cover PFTs in certain dry conditions (Nelson and Walther et al., 2024). The lack of specific Plant Functional Type (PFT) information in the EC-STILT training data may be responsible for the removal of these strong sources. The large model spread in the tropics (Fig. 4D) across the

Amazon Basin, the Congo Basin, and Oceania is the major source of global model spread in Fig. 3A.

4.2   Regional model performance

The atmospheric constraint has a strong impact on the EC-STILT estimate of regional fluxes. Similar to the global results, EC-STILT estimates regional magnitudes of NEE which are largely inconsistent (Fig. 5, Table 2) when compared with the ensemble of atmospheric inversions used in Friedlingstein et al. (2023) corrected for fire using GFED4.1 fire emissions (Fig. 5, purple lines). The regional IAV is more in agreement with atmospheric inversions, producing similar or better R^2 and magnitude when compared with X-BASE regional NEE (Fig. 5, inset text, Table 2).

Regionally, there appears to be a relationship between the coverage of the eddy-covariance and atmospheric observational networks, and the impact of the atmospheric information (Table 2). For example, in Europe (Fig. 5), both the atmospheric and eddy-covariance networks are extensive. It appears that in limited cases, the atmospheric information provides some additional information on the magnitude. EC-STILT is closer to the inversion mean with an RMSE of $0.2\,PgC\,yr^{-1}$ compared with $0.37\,PgC\,yr^{-1}$ for X-BASE. However, the IAV is very strongly conditioned on the information in eddy-covariance data; the IAV for Europe from EC-STILT is still very similar to X-BASE with a R^2 of 0.86 compared with X-BASE, and 0.31 compared with the inversion mean (Table 2). The spatial pattern of summer NEE during the peak growing season (mean JJA) is very consistent between EC-STILT and X-BASE (Fig. 6) with a spatial correlation of 0.86, demonstrating the strong influence eddy-covariance information retains when used in dual-constraint. The reduced sink in the region estimated by EC-
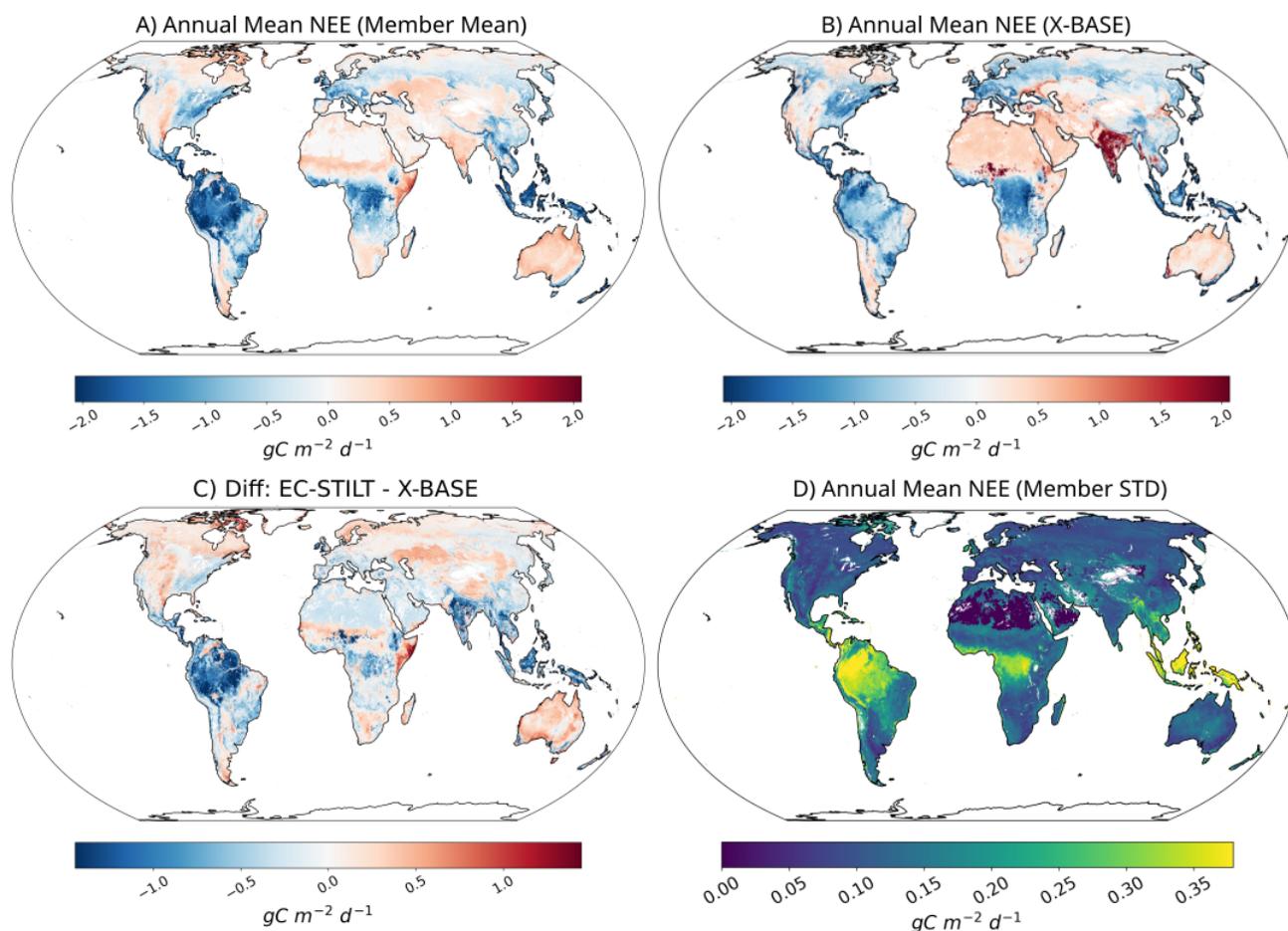
Figure 4. Spatial distribution of global annual NEE **(A)** Mean annual NEE for EC-STILT in gC m^{-2} d^{-1} **(B)** Mean annual NEE for X-BASE in gC m^{-2} d^{-1} **(C)** The difference in mean annual NEE between EC-STILT and X-BASE in gC m^{-2} d^{-1}. **(D)** The standard deviation of mean annual NEE for the EC-STILT 10-member ensemble in gC m^{-2} d^{-1}.

STILT (Fig. 5), appears in Fig. 6 to be largely an overall bias, rather than a learned difference in the spatial distribution of NEE.

In regions such as South American Tropical where there is limited eddy-covariance coverage to adequately constrain the model, additional atmospheric information does not improve the modeled NEE, and increases the model uncertainty. EC-STILT has moved away from the inversion mean, compared with X-BASE (RMSE of 3.73 PgC yr^{-1} compared with 2.05 PgC yr^{-1}). The low R^2 for EC-STILT compared with both the inversion mean and X-BASE annual time series (0.10 compared with the inversions, 0.13 compared with X-BASE) indicate that neither the eddy-covariance data nor the atmospheric signal provide sufficient constraint for EC-STILT to find similar IAV to the atmospheric inversion mean, which itself might be wrong.

The increase in the IAV for the South American Temperate region demonstrates the potential for complex effects from the atmospheric constraint. The region is only directly impacted by the ATTO tower observations in unusual mete-orological conditions, which is evident in the mean ATTO footprint in Fig. 2. The eddy-covariance record for the region is sparse, but may have similarities to other, more densely sampled regions. We see a small move away from the inversion mean compared with FLUXCOM X-BASE in the magnitude of NEE (RMSE of 0.74 PgC yr^{-1} compared with 0.59 PgC yr^{-1} for X-BASE), but an increase in the magnitude of IAV compared with FLUXCOM X-BASE (0.14 PgC yr^{-1} compared with 0.07 PgC yr^{-1}). The atmospheric constraint has added new information about this region despite very limited direct information available in training.

The atmospheric constraint also has strong effects in TransCom regions that do not directly contain or overlap the STILT regions used during training. This is because EC-STILT learns its land-surface response in environmental space of the features instead of in geographic space like an inversion. Across the globe, EC-STILT produces consistent changes over regions with similar latitude ranges (Fig. 5, right column) with a dominance of the global north in ob-
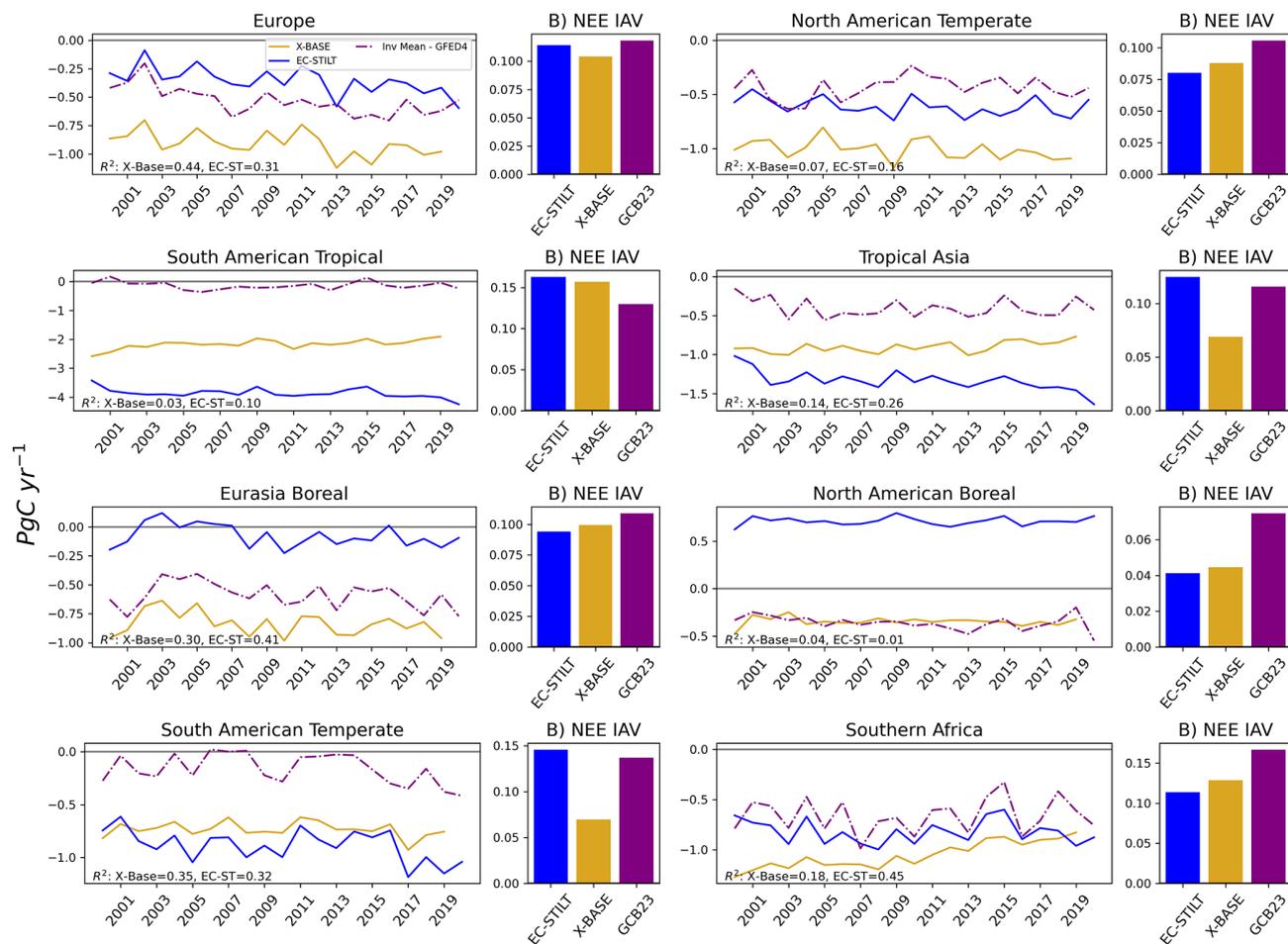
Figure 5. Annual regional NEE and IAV by TransCom region, left: the time series of the annual regional integral flux. right: the magnitude of IAV for the region. Column (**A**): Regional fluxes for TransCom regions directly impacted by atmospheric observations: Europe is impacted by observations in the HUN domain. Eurasia Boreal is impacted by observations in the ZOTTO domain. Both South American regions are directly impacted by observations in the ATTO domain. Column (**B**): Regions of similar biome and latitudinal band, by row. The blue is the ensemble mean of EC-STILT, the yellow is FLUXCOM X-BASE, the purple is the ensemble mean of the atmospheric inversions in GCB23 with GFED4.1 fire emissions removed. For each annual time series, the R^2 is reported for EC-STILT and X-BASE with regard to the fire-corrected ensemble mean

Figure 6. Mean summer NEE (JJA) in μmol CO$_2$ m^{-2} s^{-1} for EC-STILT, X-BASE, and the ensemble mean of the atmospheric inversions.

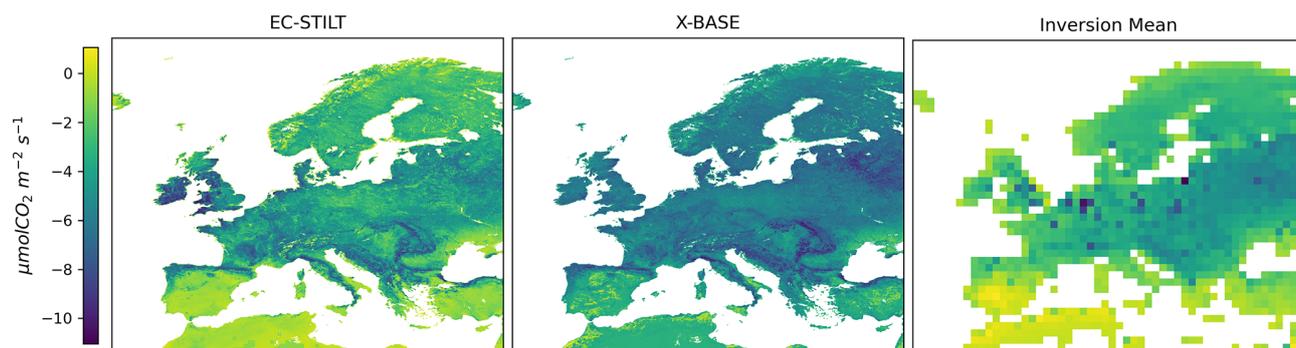**Table 2.** Comparison of different estimates with GCB: Column **(A)** shows the RMSE between the annual regional integrals in $PgC\,yr^{-1}$ of EC-STILT and X-BASE with the atmospheric inversion ensemble mean with fire removed (Invs). Column **(B)** shows the R^2 between the annual regional integrals from EC-STILT and X-BASE with regard to the atmospheric inversion ensemble mean, and between EC-STILT and X-BASE. Bold indicates the higher relative performance between EC-STILT and X-BASE for each region and metric.

	(A) Annual RMSE		**(B)** Annual R^2		
	EC-STILT, Invs	X-BASE Invs	EC-STILT Invs	X-BASE Invs	EC-STILT X-BASE
North American Boreal	1.08	**0.01**	0.01	0.01	0.39
North American Temperate	**0.21**	0.56	**0.16**	0.07	0.55
South American Tropical	3.73	**2.03**	**0.10**	0.04	0.18
South American Temperate	0.74	**0.59**	**0.32**	0.31	0.42
Northern Africa	1.13	**0.66**	0.25	**0.35**	0.05
Southern Africa	**0.20**	0.42	**0.45**	0.09	0.01
Eurasia Boreal	0.52	**0.27**	**0.41**	0.27	0.75
Eurasia Temperate	1.89	**1.56**	0.27	**0.29**	0.88
Tropical Asia	0.94	**0.51**	**0.26**	0.12	0.01
Australia	1.37	**0.42**	0.08	**0.29**	0.01
Europe	**0.21**	0.40	0.31	**0.38**	0.87

servational networks (Chu et al., 2017). The magnitude of the northern hemispheric land carbon sink is part of ongoing discrepancies in the Global Carbon Budget (O'Sullivan et al., 2024).

4.3 IAV attribution

We find that EC-STILT attributes IAV to tropical drylands in regions Aw (tropical savanah) and Bsh (semi-arid) (Fig. 7). This contrasts with the X-BASE NEE, which attributes IAV to regions Dfb (humid continental), and Dfc (sub-arctic). Both figures show the regional contribution of IAV, relative to the overall IAV of the dataset. This metric captures only the magnitude of the contribution, not the relative accuracy of the inferred IAV, which varies strongly by region (Fig. 5). Recent studies (Metz et al., 2023, 2025; Ahlström et al., 2015; Poulter et al., 2014) also suggest that arid regions are the dominant source of the IAV in global NEE. The constraints that drive this result in our system must come from atmospheric CO_2 data, and we will try to trace the source of this information in Sect. 4.5.

4.4 Eddy-covariance site-level evaluation

EC-STILT maintains good overall performance for NEE at the eddy-covariance site level, particularly at daily time scale. We use R^2 and the Nash Sutcliffe model efficiency (NSE) (Nash and Sutcliffe, 1970) to evaluate model performance with regard to the observed eddy-covariance observations, and for comparison with X-BASE. In the EC-STILT's 10-fold cross-validation for all sites (Table B4), the model achieves a median NSE of 0.66, and median R^2 of 0.77 for hourly fluxes (X-BASE: NSE 0.71, R^2 0.71), and a median NSE of 0.76 and median R^2 of 0.87 for diurnal mean NEE

(X-BASE: NSE 0.81, R^2 0.81), when comparing the observed and simulated time series. The data used for these performance metrics are the test data, held out from each cross-validation fold, and these tests are performed using the model member which did not see these data in training.

As with regional results, the site-level performance of EC-STILT appears to be strongly influenced by the level of local constraint from eddy-covariance. Figure 8 shows the hourly and daily averaged time series of NEE for three EC sites directly impacted by the atmospheric domains, one in boreal Russia (RU-Ha1), one in a German mixed forest (DE-Hai), and one Brazilian site in the Amazon basin (BR-Sa1). In the Russian site, EC-STILT (blue line) slightly improves the estimate of the mean daily cycle of the EC observations (red line) compared with X-BASE (RMSE of 0.54 compared with 0.79). When averaged to the daily scale, both EC-STILT and X-BASE produce a similar RMSE compared with the observations. EC-STILT appears to better represent the daily signal than X-BASE with a higher R^2 (0.43 compared with 0.18). EC-STILT produces a range in NEE which is slightly larger than the EC record and much greater than FLUXCOM X-BASE (yellow line). Boreal regions are only moderately constrained in EC-STILT by eddy-covariance given the large difference in annual NEE between EC-STILT and X-BASE (Fig. 5). This may allow for larger differences at the site level, as the boreal observations do not sufficiently weigh against the atmospheric information.

In the German site, DE-Hai, where the constraint from eddy-covariance is strongest, EC-STILT underestimates the hourly magnitude (Fig. 8, center left) with a mean RMSE of 0.92 for EC-STILT, 0.22 for X-BASE), but still largely captures the daily NEE, but with a larger bias than X-BASE (Fig. 8, center right) with an R^2 of 0.61 and RMSE

Figure 7. Attribution of IAV by Koppen-Geiger land cover class. The figure on the left is EC-STILT, the figure on right is X-BASE. The value represents the relative contribution of the region to the overall global IAV. The attribution is calculated by the covariance method described in (Lee et al., 2023) using annual detrended NEE anomalies.

of 2.00 compared with an R^2 of 0.83 and RMSE of 1.04 for X-BASE. In the northern extra-tropics, as seen above, the marginal impact of the atmospheric information is minimized, and EC-STILT performs most similarly with X-BASE. This mismatch in the site level results is consistent with the slightly weaker regional sink seen in the northern temperate regions.

In the Brazilian site, over the mean daily cycle EC-STILT strongly underestimates the range of NEE (RMSE of 2.52), but infers a daily value when averaged, that is comparable with X-BASE (R^2 of 0.47 for EC-STILT and X-BASE). This may indicate that the eddy-covariance and atmospheric constraints are providing confounding information about the hourly flux, which is discussed below.

4.5 Learning in feature space

We analyse in Fig. 9 the coverage of the EC-only and EC-STILT training data in terms of climate and ecological space by approximating the PDF of the two joint distributions in feature-space (TA/VPD, and EVI/NDWI). We compare the relative difference in the Kullback-Leibler distance (KLD) and Jennsen-Shannon metric (JS) between the PDF of the full, or natural distribution of a random subset of all pixels, either globally or by KG class, and the two training sets (Fig. 9, Tables B1–B3). The two variable pairs were chosen to create easily interpretable visualizations. We use subsets of the full 10-dimensional distribution to save on computational costs.

At the global scale, the distribution of the EC training set is more concentrated in cooler, and moderately productive regions than the EC-STILT set (Fig. 9, top panel, compare the higher density in the warmer, more productive regions). In the second row, we can see specifically how the training data changes the representation of a region which is not directly observed. As with the global distributions, the EC-STILT distribution covers the warmer, more water-stressed regions (Fig. 9, middle panel TA/VPD) and in the wetter, more highly vegetated regions (Fig. 9, middle panel EVI/NDWI).

This distributional approach also explains why the model under-performs in regions that are directly constrained by the atmospheric towers. In the third row, the Köppen-Geiger class Cfb (Temperate Oceanic) which covers most of western Europe, we can see that in both variable sets, the EC-only distribution is closer to the natural distribution (Fig. 9, bottom row). Because of the environmental learning of EC-STILT, this means that the inclusion of atmospheric towers may reduce the model's skill in this region.

The specific impact of individual towers, and their interactions in feature space, are inconsistent across regions, variable pairs, and metrics. In Tables B1 to B3, we expand the analysis by quantifying the effect of adding or removing individual atmospheric towers from training. The optimal distribution, shown in bold, characterized by the minimum distance between the two PDFs, is enriched by adding 1–3 towers in 49 of 52 regions across both distributions (Tables B1, B2). However when the full distribution from all towers are considered (Table B3), the improvement of the tower-enriched distributions is more modest, with only 9 of
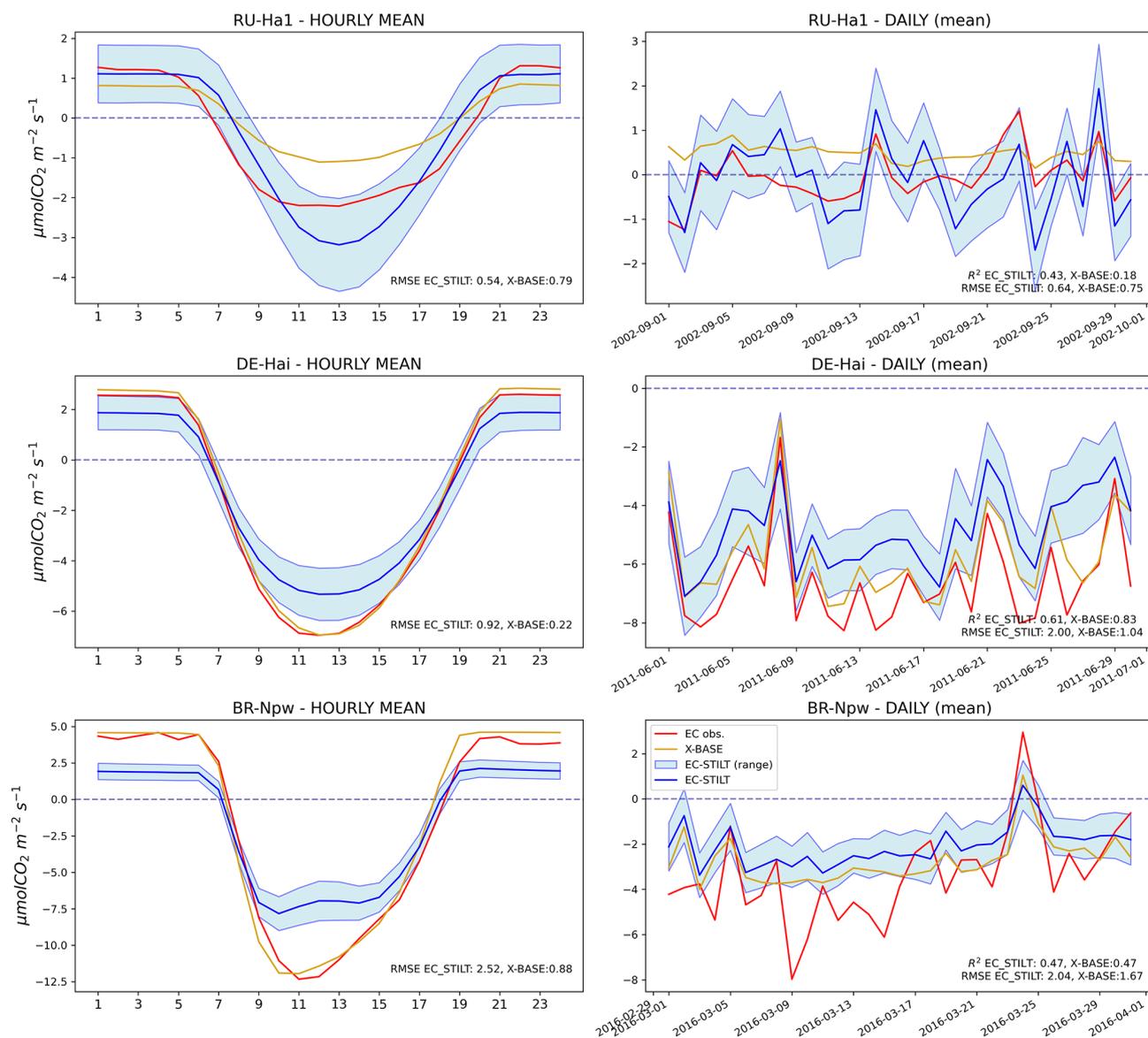
Figure 8. EC tower results across latitude bands: Each row is a tower. The left column is the mean hourly NEE, the right column is the daily mean across an example month. The blue line is the member mean of EC-STILT, the yellow line is the X-BASE NEE, and the red dotted line is the eddy-covariance observations for the tower. All values are in $\mu mol\,C\,m^{-2}\,s^{-1}$.

26 regions (including global) having improved coverage by the towers in both variable pairs.

5 Discussion

The core logic of our model is a synergistic sharing of information between the bottom-up and top-down constraints. The key innovation in our system is the combination of the atmospheric term with the ecosystem-level objective term derived from eddy-covariance data, and their joint projection onto a space that is not the "traditional" geographic distribution of NEE. Instead, EC-STILT learns in "environmental response" space, or a set of relations between the driver datasets (features) and the output NEE. The STILT Atmospheric transport model provides the spatiotemporal locations where ecosystem NEE variations contribute to a downwind CO$_2$ observation. This large number of locations over the near-field can be modeled and their output constrained by the atmospheric observation. Considered independently this is underconstrained; one atmospheric observation is insufficient to constrain the multiple modeled NEE estimates under the STILT footprint. This lack of constraint is also experienced in many inverse modeling problems using atmospheric data (see Peylin et al., 2013; Gaubert et al., 2023 for an overview). However, because the "environmental response" space of the model is conditioned by the ecosystem-
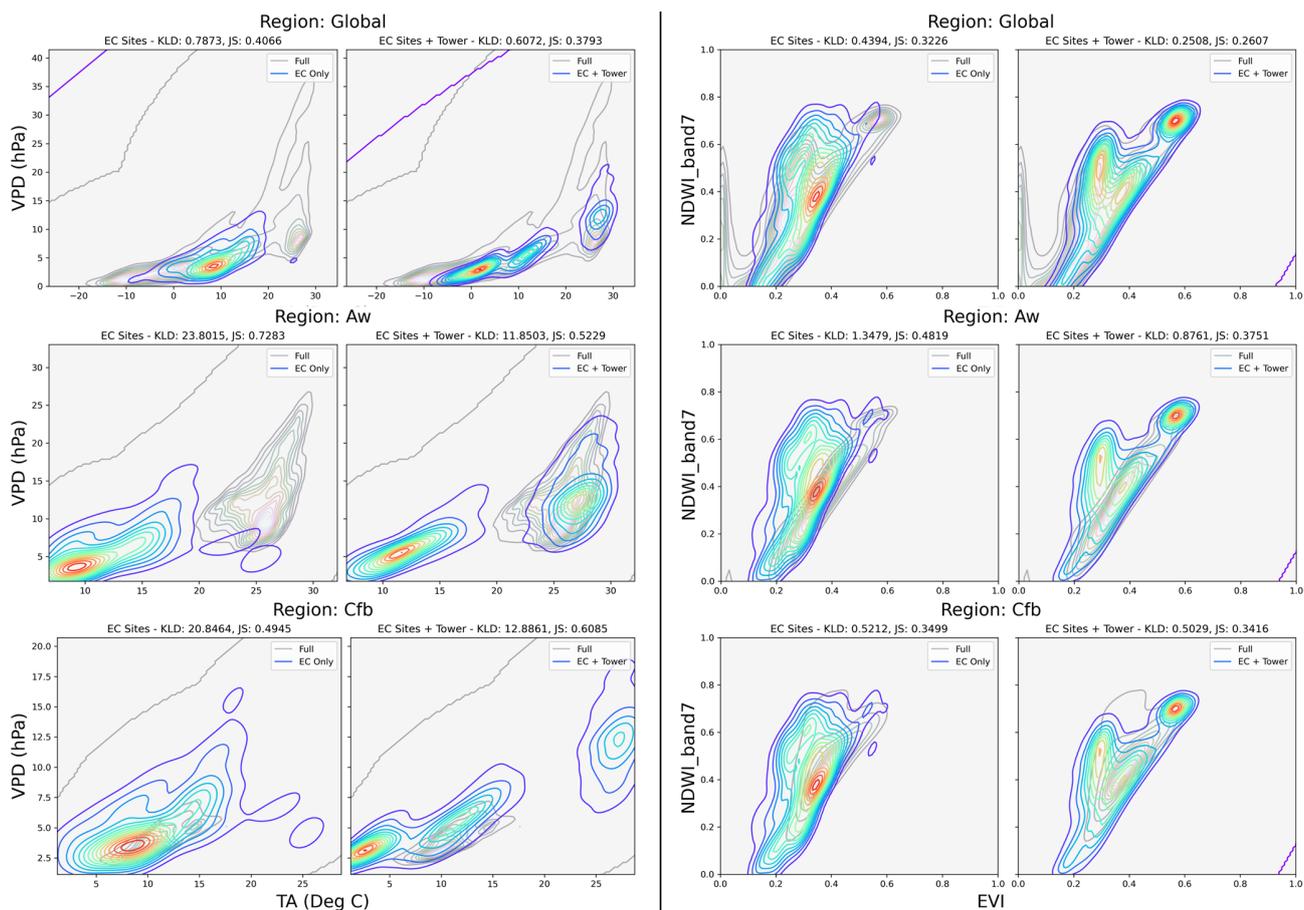
Figure 9. Comparison between training set distributions in two multivariate spaces. Both columns show the difference in distribution between the EC-only training set (left), and the EC+Towers training set (right), and how they compare with the full distribution of the global or regional data in feature space. The left is TA and VPD, the right is EVI and NDWI. The top row is the full global set, the middle row is Köppen-Geiger class Aw (Tropical Dry Winter), the bottom row is Köppen-Geiger class Cfb (Temperate Oceanic). The contour lines represent the probability distribution function (PDF) of each distribution. Two metrics are calculated for each: Kullback-Leibler (KLD) and Jennsen-Shannon divergence (JS), which are measures of distance between the two PDFs.

level objective term, the response across the STILT footprint is tightly bound to the local, better-constrained eddy-covariance record. Learning is thus jointly informed by both data sources while maintaining, outside of the tropics, the local daily performance of bottom-up models like FLUXCOM V1 (Jung et al., 2020) and X-BASE (Nelson and Walther et al., 2024).

EC-STILT efficiently learns driver/NEE relationships globally which adds new information on the regional and global dynamics of NEE. This is demonstrated by the improved representation of IAV of global NEE (Fig. 3) and regional NEE (Fig. 5). There is no explicit representation in the training data of the atmospheric IAV; the atmospheric constraint is only present for three regional domains (around the ZOTTO, ATTO, and HUN towers). These domains were chosen to represent a tropical (ATTO), extra-tropical (HUN), and boreal (ZOTTO) eco-region. Each tower has observations across a single year (2019 for ZOTTO and Hegyhát-

sál, 2015 for ATTO) chosen for reasons of data availability. These relatively small samples of the available atmospheric information are sufficient to partially embed the atmospheric signal in the ecosystem response of EC-STILT, and modify the spatial and temporal distributions of estimated NEE.

We selected a limited subset of towers and training years to balance computational cost and model performance in this proof-of-concept system. The computational cost of additional towers and years is primarily associated with creating STILT footprints for a new temporal and spatial domain, which is costly both in terms of computational and human effort. We selected three sites with the specific goal to achieve good coverage of different geographic, climate and ecological zones. This precluded the use of other towers which we believe might have provided valuable constraints in training. The NOAA ObsPack data (Masarie et al., 2014) could provide a large volume of additional training data, both towers and years. Following the analysis of Sect. 4.5 we find that

given the structure of our model, one might be able to increase the performance of the model with a targeted selection of towers. These would optimally represent the natural distribution of the land surface, rather than including all measurement towers across observational networks. Future experiments, similar to the EC representation analysis in Pallandt et al. (2022), can be performed to identify additional towers or years that might yield largest improvements in our predictions.

There is no explicit representation in the data of the atmospheric growth rate, nor mass-balance, yet these are reproduced to an impressive degree. The atmospheric observations are regional and instant, with no direct information on the long-term state of the atmosphere. Because the atmospheric transport system is in fact not run during the calculation of global NEE, there is therefore no atmospheric information available to EC-STILT during the creation of our global NEE data. The potentially realistic global integral and spatial distribution (Fig. 3) demonstrates that the model is able to learn a consistent ecosystem-level response globally from this sparse set of local observations with an improved representation of global dynamics. Of particular note is that these limited observations improve the estimate of IAV across the entire period of analysis, and thus learned relations hold under a range of environmental conditions in each pixel.

Although subject to large uncertainty, the magnitude of the annual flux has moved closer to the GCB23 atmospheric inversion estimate of global NEE. The global magnitude is strongly conditioned, however on the inclusion and handling of the non-biospheric flux data during training; as the total amount of NBF in a region increases, by necessity, the inferred NEE must decrease to keep the local instantaneous equation balanced (Eq. 1). The total contributions to the well-mixed atmosphere over time are complex (Ciais et al., 2022), and often not available in spatially explicit data products. The current study used fire and fossil fuel emissions, but other terms such as riverine fluxes (see Botía et al., 2025, for a discussion at the ATTO tower) may play strong local or global roles in the regional CO$_2$ balance, and should be included in the tower loss calculation.

Because EC-STILT has no direct access to long-term information about the state of the atmosphere, and yet produces an estimate of global annual NEE which is closer to the atmospheric inversions, there must be some information within the available training data which includes this information. The identically constructed model, trained without the inclusion of the STILT operator (Sect. 3.6) does not produce an increased IAV (see Fig. B1), which removes the influence of the model architecture. We further evaluate the influence on IAV of the LBC (Eq. 1), which is derived from atmospheric inversions. In the results from our mean-LBC model, this change did not reduce the observed IAV (see Fig. B2), which means the LBC is not responsible for the improvements in the IAV in the standard EC-STILT.

To understand the increased IAV in the tropical dry regions (see Sect. 4.3), we hypothesize that our model can better represent the local NEE responses to climate variations in these regions through a more complete representation of the natural distribution of the biophysical drivers of NEE, as shown in Fig. 9. This improvement over the information provided by the EC towers in feature-space explains EC-STILT's improved long-term performance using only a single year of observations. EC-STILT learns to represent NEE using only 10 variables at an hourly time-step. With no static variables, such as latitude, elevation, or PFT information, the model can be considered largely independent of a particular spatial domain. The model only learns to map from a training distribution of drivers to a training distribution of NEE. The quality and completeness of this training distribution determines the models capacity to capture the global phenomena.

The results from Upton et al. (2024) also support the distribution hypothesis presented above. In that study the atmospheric constraint uses a limited number of fixed pixel locations to infer regional totals. This created an NEE product which was much closer in magnitude and seasonality to an ensemble of atmospheric inversions, but did not improve the IAV. This version of atmospheric constraint does not fundamentally change the model's available view of the land surface. From this we can see that the inclusion of training data which fully includes the IAV signal, but that does not improve the distributional representation of regions from which the IAV emerges from the local variance, does not improve the model's ability to capture the IAV.

In Fig. 3A, it appears that the atmospheric signal increases the uncertainty around the magnitude and distribution of tropical NEE, when compared with X-BASE. The ensemble members show a large spread across tropical regions, and at the ecosystem-level, the ensemble mean has difficulty reproducing the magnitude of hourly observations from the eddy-covariance record (Fig. 8). EC-STILT's tropical shortcomings may be a combination of several structural factors; model type, potential mismatches in the atmospheric operator, and limited observational support, which are discussed below.

When compared with X-BASE, the core model structure is important; X-BASE is a XGBoost regressor, which have been demonstrated to be robust in this domain compared with neural networks which have more capacity to overfit (Nelson and Walther et al., 2024; Kraft et al., 2025; Jung et al., 2020). Because our objective function is sensitive to ecosystem and regional results across all three towers, it is possible that within the overall learning process, some local or regional overfitting is occurring. We attempt to reduce overfitting using early stopping of training according to the validation accuracy, and in the model using batch normalization. Potentially, this overfitting could lead to large intra-model uncertainty between neural network members compared with an XGBoost ensemble. Current implementations of XGBoost in the Python programming language (compati-

ble with the FLUXCOM-X framework) were not considered, because they do not allow for complex objective functions with multiple terms, sensitive to different spatial scales. Any additional experimentation with different model structures was outside of the scope of this study.

The tropical response from EC-STILT is also confounded by a specific effect in the performance of the STILT atmospheric operator during training. For HUN and ZOTTO, EC-STILT achieved high Pearson's R values (0.65–0.8) for the inferred and observed land contribution to the CO$_2$. However for ATTO, although the RMSE during training between the inferred and observed land contribution was similar to the training RMSE for HUN and ZOTTO, Pearson's R was near zero (0.0–0.15, see Fig. A5). This means that the system was able to infer the correct magnitude of the contribution, but unable to get the signal correctly located in time. This poor temporal registration may be attributed to a missing NBF contribution to the signal which has a strong temporal component, such as the mentioned riverine evasion of CO$_2$ (Botía et al., 2025). Or the issue may be with the transport model or its meteorological forcing. Our STILT footprints are calculated at 0.25° spatial resolution. This resolution was chosen because of computational constraints in our training. It is possible that this spatial resolution does not adequately capture some aspect of the atmospheric processes in the tropics. In relation to the relatively small contribution of intra-model uncertainty within LPDMs (Hegarty et al., 2013), meteorological uncertainty is a more important term in the overall transport uncertainty (Angevine et al., 2014). However, it was out of scope for the current work to estimate that uncertainty with regard to the output STILT sensitivities and and resulting concentration estimates. Lastly, the temporal issue at ATTO may have to do with the rate of air exchange through the canopy. This could introduce a time lag into the release of CO$_2$ out of the canopy and into the free atmosphere (Faassen et al., 2024).

Both bottom-up (Chu et al., 2017) and top-down (Palmer et al., 2019) observational records are limited and uncertain in the tropics. This limited record forces a data-driven model to extrapolate its response for a range of missing or under-sampled parts of the tropical distribution of the driver variables. It is unrealistic for a model constrained only by the existing observational record to move meaningfully beyond previous work on estimating tropical fluxes. Although the boreal region around ZOTTO is also subject to the same uncertainties with regard to the observational record, the uncertainty relative to the magnitude of the flux does not impact global results to the same extent.

In the eddy-covariance site results (Fig. 8) we may see an example of atmospheric information confounding and augmenting the eddy-covariance signal. Both boreal and tropical regions are less well constrained by eddy-covariance than the northern extra-tropics, which increases the marginal impact of the atmospheric constraint. EC-STILT was able to reproduce both the Russian site's (RU-Ha1) eddy-covariance

hourly time series and the ZOTTO time series with high fidelity. Indeed, EC-STILT estimates an hourly variability which is higher than the hourly eddy-covariance data. In the Brazilian site (BR-Npw), where EC-STILT was unable to reproduce the ATTO time series, we see the eddy-covariance finding a low-variability estimate of the daily mean. This may represent confounding between constraints.

While the absolute increase in the R^2 of IAV with regard to the GCB23 inversions is modest (0.42 for EC-STILT, 0.02 for X-BASE), it represents a meaningful increase over previous data-driven flux models (Jung et al., 2019; Nelson and Walther et al., 2024). As seen above in Sect. 4.2 and Fig. 7, EC-STILT improves the estimation of IAV in regions in the Southern Hemisphere which are known to contribute to a large fraction of the global IAV. However EC-STILT fails to improve the representation of IAV in the Northern Hemisphere, or where the EC observational record dominates. Therefore, the modest gain in R^2, can be seen as a meaningful gain in the representation of the land surface in regions which are otherwise poorly represented in the EC record.

5.1 Future work

This study finds improvement in a number of important metrics in a bottom-up NEE model that is additionally constrained by three atmospheric tall-towers over one year. As noted above, neither the eddy-covariance record nor the atmospheric observational record have strongly representative global coverage, and additional atmospheric information over new geographic regions could be a benefit. Therefore an important future development of a dual-constrained model is the inclusion of additional tower and non-tower data streams: Additional tall towers could provide stronger constraint in well-measured regions. Non-tower data could be leveraged to provide information on regions that are poorly covered by the current observational network. The current model could run, as-is, against aircraft observations. With a modification of the STILT transport system in the objective function, the structure of this study would allow a bottom-up model to be constrained by satellite retrievals of total column CO$_2$ (XCO$_2$). XCO$_2$ observations could be selected according to their potential to include biomes of specific interest in their upwind footprints. This would enrich our capacity to model NEE across un- and under-sampled biomes. This extension into XCO$_2$ introduces several possible issues: to model the full column of CO$_2$, multiple STILT footprints must be run for every observation at different elevations (Wu et al., 2018), and modeled fluxes transported multiple times. This may be a technical challenge to build effective training systems. The inclusion of the full column CO$_2$ may dilute the available information on the biogenic flux in the observation, reducing the information gain to the bottom-up model.

The performance of EC-STILT in the tropics show the limitations of adding atmospheric information without adequate support in the eddy-covariance record. In the harmo-

nized model described above, the ecosystem-level information appears to provide the "backbone" of the model's performance. Future modeling efforts will benefit from the ongoing efforts to improve and enlarge the corpus of available eddy-covariance sites and data, such as a potential extension of the FLUXNET 2015 dataset (Pastorello et al., 2020). As the eddy-covariance record better describes and constrains key biomes, additional atmospheric constraint may be able to provide improved information on the larger dynamics of regional and global NEE, while reducing the uncertainty.

An important aspect of transport modeling which is not addressed in the current study is the impact of model and meteorological uncertainty. In our analysis, we treat the output of the STILT model as definitive, relative to the meteorological fields in ECMWF short-term forecast. Computational schemes for addressing model uncertainty, multi-LPDM ensembling, or in-model ensembling for STILT by sampling within the stochastic particles could be attempted, without a major increase in computational and I/O workload. In this study, STILT runs are forced with ECMWF short-term forecast data (Botia et al., 2022), and the model is trained using ERA5 reanalysis data. Using identical meteorological data for STILT and the model drivers may also reduce uncertainty.

Conceptually, the current work accounts for two major terms in the non-biogenic flux budget; fires and fossil fuel emissions. In the Amazon and in boreal Russia, fire is the largest and the most relevant term, as the term accounts for both fire, and also for part of the instantaneous flux from land-cover/land-use change associated with biomass burning (Cochrane and Laurance, 2008). In the European domain, fossil fuels are the dominant non-biogenic flux signal. As noted in the Sect. 2, NEE at the eddy-covariance tower and atmospheric CO$_2$ are not directly comparable. The dual-constraint in this study is created using the residual of an observation and non-biogenic fluxes. Any biogenic or non-biogenic flux terms which are not included, or adequately represented temporally and spatially will bias the atmospheric target that the model is trying to match in training. An example of this would be the non-fire disturbance fluxes, or regrowth after disturbance. Another potentially important term is the instantaneous riverine flux of CO$_2$, coming from lateral riverine transport. These terms could push the relative carbon balance towards CO$_2$ release (disturbances) or towards CO$_2$ assimilation (regrowth), and during retraining the model would attempt to match this new local target balance. As seen above, because of the distributional nature of the model's learning and inference, this new local balance could modify the global NEE response. This reduces the effectiveness of the technique. Any additional spatially explicit NBF data may therefore, improve this comparison, and increase the fidelity of the overall magnitude and may reduce the uncertainty in the global result.

6 Conclusions

This study demonstrates a novel method for combining bottom-up and top-down information in the objective function of a data-driven flux model. We wanted to test the effectiveness of a physics-based operator that transports modeled flux density into atmospheric concentration. And we wanted to determine whether a temporally and geographically-sparse constraint could produce valid global results. The EC-STILT model in this study is still strongly conditioned by the bottom-up constraint from eddy-covariance, and so maintains the strengths of the overall bottom-up approach, but now with access to information on land surface far beyond the eddy-covariance network.

A classical top-down atmospheric inversion matches the observed atmospheric signal but lacks the framework to accurately distribute that signal down into geographic/biophysical space. Bottom-up models can translate EC measurements into their latent space, but do not have any additional information about how inferred fluxes influence the global signal. In contrast, our hybrid system in this study can take information from the atmospheric observations and accurately distribute it into the learned latent space of the model. Because this latent space still contains information from the eddy-covariance record, this allows for a valid mapping down from the atmosphere to the model, and from the model to an estimate of NEE at multiple scales which is comparable both with both atmospheric and ecosystem-level observations. Using our system, a dual-constrained model sits directly between bottom-up and top-down approaches, directly addressing their limitations while maintaining their advantages.

Appendix A: Data

Table A1. Citation data for the 294 sites used in training.

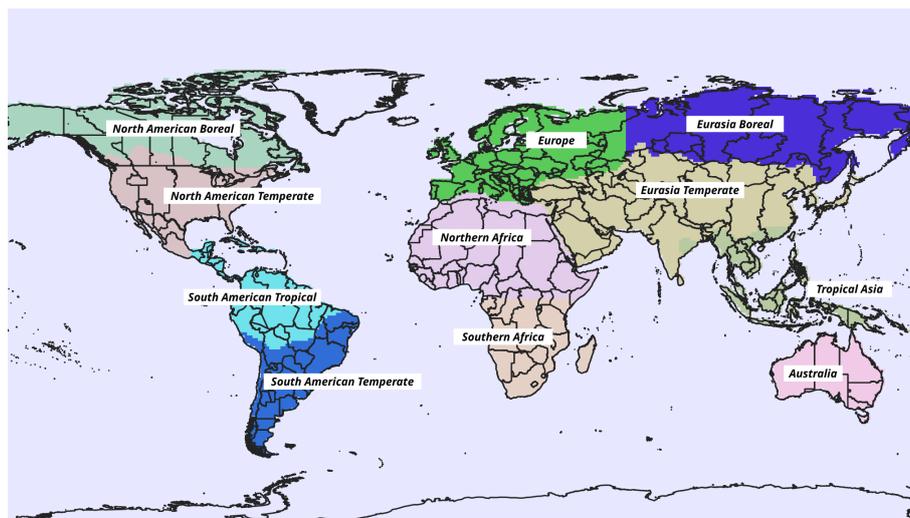|					
AR-SLu (Garcia et al., 2016)	AR-TF1 (Kutzbach, 2021)	AR-Vir (Posse et al., 2016)	AT-Neu (Wohlfahrt et al., 2016)	AU-ASM (Cleverly and Eamus, 2016b)	AU-Ade (Beringer and Hutley, 2016c)
AU-Cpr (Meyer et al., 2016)	AU-Cum (Pendall and Griebel, 2016)	AU-DaP (Beringer and Hutley, 2016b)	AU-DaS (Beringer and Hutley, 2016f)	AU-Dry (Beringer and Hutley, 2016e)	AU-Emr (Schroder et al., 2016)
AU-Fog (Beringer and Hutley, 2016a)	AU-Gin (Macfarlane et al., 2016)	AU-RDF (Beringer and Hutley, 2016d)	AU-Rob (Liddell, 2016)	AU-TTE (Cleverly and Eamus, 2016a)	AU-Tum (Woodgate et al., 2016)
AU-Wac (Beringer et al., 2016b)	AU-Whr (Beringer et al., 2016a)	AU-Wom (Arndt et al., 2016)	AU-Ync (Beringer and Walker, 2016)	BE-Bra (Team and Centre, 2022b)	BE-Dor (Team and Centre, 2022b)
BE-Lcr (RI, 2021a)	BE-Lon (Team and Centre, 2022b)	BE-Maa (Team and Centre, 2022b)	BE-Vie (Team and Centre, 2022b)	BR-Npw (Vourlitis et al., 2022)	BR-Sal (Saleska, 2016)
BR-Sa3 (Goulden, 2016d)	CA-Cbo (Staebler, 2022)	CA-DB2 (Knox, 2022)	CA-DBB (Christen and Knox, 2022)	CA-ER1 (Wagner-Riddle, 2021)	CA-Gro (McCaughey, 2016)
CA-LP1 (Black, 2021)	CA-Man (Amiro, 2016b)	CA-NS2 (Goulden, 2016a)	CA-NS3 (Goulden, 2016b)	CA-NS4 (Goulden, 2016c)	CA-NS5 (Goulden, 2016g)
CA-NS6 (Goulden, 2016e)	CA-NS7 (Goulden, 2016f)	CA-Oas (Black, 2016b)	CA-Obs (Black, 2016a)	CA-Qfo (Margolis, 2016)	CA-SF1 (Amiro, 2016c)
CA-SF2 (Amiro, 2016a)	CA-SF3 (Amiro, 2016d)	CA-TP1 (Arain, 2016b)	CA-TP2 (Arain, 2016a)	CA-TP3 (Arain, 2022b)	CA-TP4 (Arain, 2016c)
CA-TPD (Arain, 2022a)	CG-Tch (Nouvellon, 2016)	CH-Aws (Team and Centre, 2022b)	CH-Cha (Team and Centre, 2022b)	CH-Dav (Team and Centre, 2022b)	CH-Fru (Team and Centre, 2022b)
CH-Lae (Team and Centre, 2022b)	CH-Oe1 (Ammann, 2016)	CH-Oe2 (Team and Centre, 2022b)	CN-Cha (Zhang and Han, 2016)	CN-Cng (Dong, 2016)	CN-Dan (Shi et al., 2016)
CN-Din (Zhou and Yan, 2016)	CN-Du2 (Chen, 2016k)	CN-Du3 (Shao, 2016b)	CN-HaM (Tang et al., 2016)	CN-Qia (Wang and Fu, 2016)	CN-Sw2 (Shao, 2016a)
CZ-BK1 (Team and Centre, 2022b)	CZ-KrP (Team and Centre, 2022b)	CZ-Lnz (Team and Centre, 2022b)	CZ-Lnz (Team and Centre, 2022b)	CZ-RAJ (Team and Centre, 2022b)	CZ-Stn (Team and Centre, 2022b)
CZ-wet (Team and Centre, 2022b)	DE-Akm (Team and Centre, 2022b)	DE-Geb (Team and Centre, 2022b)	DE-Gri (Team and Centre, 2022b)	DE-Hai (Team and Centre, 2022b)	DE-HoH (Team and Centre, 2022b)
DE-Hte (Team and Centre, 2022b)	DE-Hzd (Team and Centre, 2022b)	DE-Kli (Team and Centre, 2022b)	DE-Lkb (Lindauer et al., 2016)	DE-Lnf (Knohl et al., 2016)	DE-Obe (Team and Centre, 2022b)
DE-RuR (RI, 2022b)	DE-RuW (Team and Centre, 2022b)	DE-RuW (Team and Centre, 2022b)	DE-Seh (Schneider and Schmidt, 2016)	DE-SfN (Klatt et al., 2016)	DE-Spw (Bernhofer et al., 2016)
DE-Tha (Team and Centre, 2022b)	DE-Zrk (Sachs et al., 2016)	DK-Eng (Pilegaard and Ibrom, 2016)	DK-Fou (Olesen, 2016)	DK-Gds (RI, 2022b)	DK-Sor (Team and Centre, 2022b)
ES-Abr (Team and Centre, 2022b)	ES-Agu (Team and Centre, 2022b)	ES-Amo (Poveda et al., 2016)	ES-Cnd (Team and Centre, 2022b)	ES-LJu (Team and Centre, 2022b)	ES-LM1 (Team and Centre, 2022b)
ES-LM2 (Team and Centre, 2022b)	ES-LgS (Reverter et al., 2016b)	ES-Ln2 (Reverter et al., 2016a)	FI-Hyy (Team and Centre, 2022b)	FI-Jok (Lohila et al., 2016)	FI-Ken (Team and Centre, 2022b)
FI-Let (Team and Centre, 2022b)	FI-Lom (Aurela et al., 2016a)	FI-Qvd (Team and Centre, 2022b)	FI-Sii (Team and Centre, 2022b)	FI-Sod (Aurela et al., 2016b)	FI-Var (RI, 2022b)
FR-Aur (Team and Centre, 2022b)	FR-Bil (Team and Centre, 2022b)	FR-EM2 (RI, 2022b)	FR-FBn (Team and Centre, 2022b)	FR-Fon (Team and Centre, 2022b)	FR-Gri (Team and Centre, 2022b)
FR-Hes (Team and Centre, 2022b)	FR-LBr (Berbigier and Loustau, 2016)	FR-LGt (RI, 2022b)	FR-Lam (Team and Centre, 2022b)	FR-Pue (Ourcival, 2016)	FR-Tou (RI, 2022b)
GF-Guy (Team and Centre, 2022b)	GH-Ank (Valentini et al., 2016b)	GL-Dsk (RI, 2022b)	GL-NuF (Hansen, 2016)	GL-ZaH (Lund et al., 2016b)	GL-ZaF (Lund et al., 2016a)
IE-Cra (Team and Centre, 2022b)	IL-Yat (Team and Centre, 2022b)	IT-BCi (Team and Centre, 2022b)	IT-BFt (RI, 2022b)	IT-CA1 (Sabbatini et al., 2016c)	IT-CA2 (Sabbatini et al., 2016a)
IT-CA3 (Sabbatini et al., 2016b)	IT-Col (Matteucci, 2016)	IT-Cp2 (Team and Centre, 2022b)	IT-Cpz (Valentini et al., 2016a)	IT-Isp (Gruening et al., 2016b)	IT-La2 (Cescatti et al., 2016)
IT-Lav (Team and Centre, 2022b)	IT-Lsn (RI, 2022b)	IT-MBo (Team and Centre, 2022b)	IT-Noe (Spano et al., 2016)	IT-PT1 (Manca and Goded, 2016)	IT-Ren (Team and Centre, 2022b)
IT-Ro1 (Valentini et al., 2016c)	IT-Ro2 (Papale et al., 2016)	IT-SR2 (Gruening et al., 2016a)	IT-SRo (Gruening et al., 2016a)	IT-Tor (Team and Centre, 2022b)	JP-MBF (Kotani, 2016)
JP-SMF (Kotani)	MX-Tes (Yepez and Garatuza, 2021)	MY-PSO (Kosugi and Takanashi, 2016)	NL-Hor (Dolman et al., 2016a)	NL-Loo (Team and Centre, 2020a)	PA-SPn (Wolf et al., 2016b)
PA-SPs (Wolf et al., 2016a)	PE-QFR (Griffis and Roman, 2021)	RU-Che (Merbold et al., 2016)	RU-Cok (Dolman et al., 2016b)	RU-Fy2 (Team and Centre, 2022b)	RU-Fyo (Team and Centre, 2022b)
RU-Hal (Belelli et al., 2016)	SD-Dem (Ardö et al., 2016)	SE-Deg (Team and Centre, 2022b)	SE-Htm (Team and Centre, 2022b)	SE-Lnn (Team and Centre, 2020a)	SE-Nor (Team and Centre, 2022b)
SE-Ros (Team and Centre, 2022b)	SE-Svb (Team and Centre, 2022b)	SJ-Adv (Christensen, 2016)	SJ-Blv (Boike et al., 2016)	SN-Dhr (Tagesson et al., 2016)	US-A32 (Billesbach et al., 2022)
US-AR1 (Billesbach et al., 2016b)	US-AR2 (Billesbach et al., 2016a)	US-ARM (Biraud et al., 2022)	US-ARb (Torn, 2016b)	US-ARc (Torn, 2016a)	US-Bi2 (Rey-Sanchez et al., 2022a)
US-BZB (Euskirchen, 2022b)	US-BZF (Euskirchen, 2022c)	US-BZS (Euskirchen, 2022d)	US-BZo (Euskirchen, 2022a)	US-Bi1 (Rey-Sanchez et al., 2022b)	US-CRT (Chen and Chu, 2016b)
US-Bio (Goldstein, 2016)	US-CF1 (Huggins, 2021)	US-CF2 (Huggins, 2022c)	US-CF3 (Huggins, 2022a)	US-CF4 (Huggins, 2022b)	US-EDN (Oikawa, 2021)
US-CS1 (Desai, 2022a)	US-CS2 (Desai, 2022c)	US-CS3 (Desai, 2022d)	US-CS4 (Desai, 2022b)	US-Cop (Bowling, 2016)	US-Ha1 (Munger, 2016)
US-GBT (Massman, 2016)	US-GLE (Massman, 2016)	US-Goo (Meyers, 2016b)	US-HB1 (Forsythe et al., 2021)	US-HWB (Goslee, 2022)	US-Ivo (Zona and Oechel, 2016b)
US-Hn3 (Liu et al., 2022)	US-Ho2 (Hollinger, 2022)	US-IB2 (Matamala, 2016)	US-ICs (Euskirchen et al., 2022a)	US-IC1 (Euskirchen et al., 2022b)	US-KS3 (Hinkle, 2022)
US-Jo2 (Vivoni and Perez-Ruiz, 2022)	US-KFS (Brunsell, 2022a)	US-KLS (Brunsell, 2022b)	US-KS1 (Drake and Hinkle, 2016a)	US-KS2 (Drake and Hinkle, 2016b)	US-Me1 (Law, 2016c)
US-LWW (Meyers, 2016a)	US-Lin (Fares, 2016)	US-Los (Desai, 2016c)	US-MMS (Novick and Phillips, 2022)	US-MOz (Wood and Gu, 2022)	US-Mpj (Litvak, 2021)
US-Me2 (Law, 2022)	US-Me3 (Law, 2016a)	US-Me4 (Law, 2016e)	US-Me5 (Law, 2016d)	US-Me6 (Law, 2016b)	US-Ne3 (Suyker, 2016a)
US-Myb (Sturtevant et al., 2016)	US-NGB (Torn and Dengel, 2021)	US-NR1 (Blanken et al., 2022)	US-Ne1 (Suyker, 2022)	US-Ne2 (Suyker, 2016b)	US-Prr (Kobayashi and Suzuki, 2016)
US-ONA (Silveira, 2021)	US-ORv (Bohrer, 2021)	US-OWC (Bohrer and Kerns, 2022)	US-Oho (Chen et al., 2016)	US-PFa (Desai, 2016d)	US-Rwe (Flerchinger and Reba, 2022)
US-Rms (Flerchinger, 2022c)	US-Ro1 (Baker et al., 2022)	US-Ro4 (Baker and Griffis, 2022a)	US-Ro5 (Baker and Griffis, 2021)	US-Ro6 (Baker and Griffis, 2022b)	US-Sne (Shortt et al., 2022)
US-Rwf (Flerchinger, 2022a)	US-Rws (Flerchinger, 2022b)	US-SRC (Kurc, 2022)	US-SRG (Scott, 2016a)	US-SRM (Scott, 2016b)	US-Tw2 (Sturtevant et al., 2022)
US-Snf (Kusak et al., 2022)	US-Sta (Ewers and Pendall, 2016)	US-Syv (Desai, 2016b)	US-Ton (Baldocchi and Ma, 2016)	US-Tw1 (Valach et al., 2021)	US-UMB (Gough et al., 2016)
US-Tw3 (Chamberlain et al., 2022)	US-Tw4 (Sanchez et al., 2016)	US-Tw5 (Valach et al., 2022)	US-Twt (Baldocchi, 2016)	US-UM3 (Bohrer, 2022)	US-Wi0 (Chen, 2016g)
US-UMd (Gough et al., 2022)	US-Var (Baldocchi et al., 2016)	US-WCr (Desai, 2016a)	US-WPT (Chen and Chu, 2016a)	US-Whs (Scott, 2016d)	US-Wi6 (Chen, 2016h)
US-Wi1 (Chen, 2016e)	US-Wi2 (Chen, 2016j)	US-Wi3 (Chen, 2016b)	US-Wi4 (Chen, 2016d)	US-Wi5 (Chen, 2016a)	US-xBR (Network, 2022)
US-Wi7 (Chen, 2016i)	US-Wi8 (Chen, 2016f)	US-Wi9 (Chen, 2016f)	US-Wjs (Litvak, 2022)	US-Wkg (Scott, 2016c)	

Figure A1. TransCom land regions.
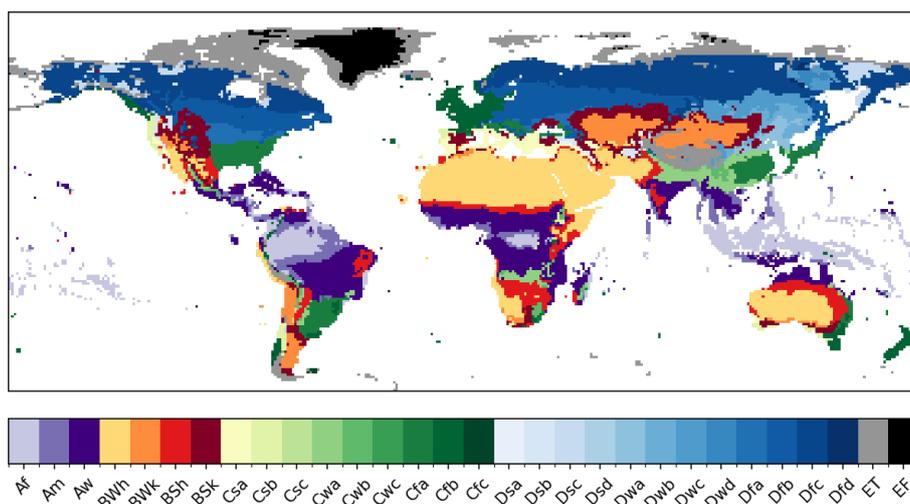
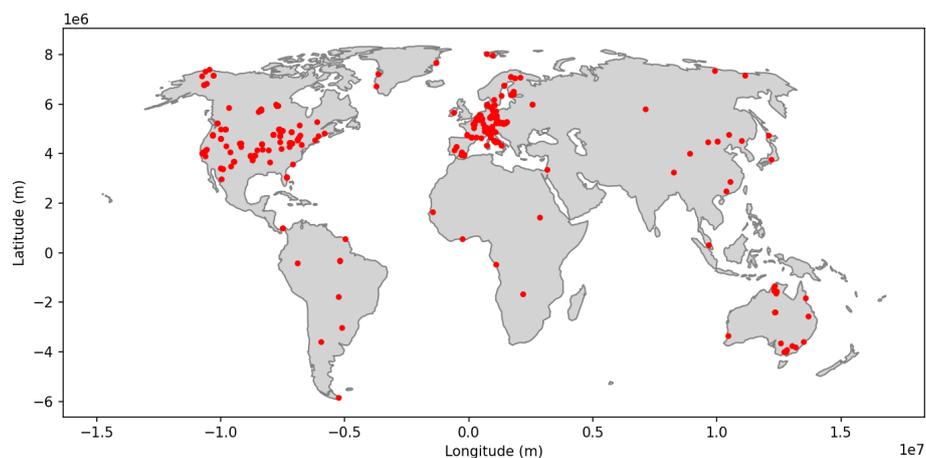**Figure A2.** Köppen-Geiger (KG) classification system from Beck et al. (2018).

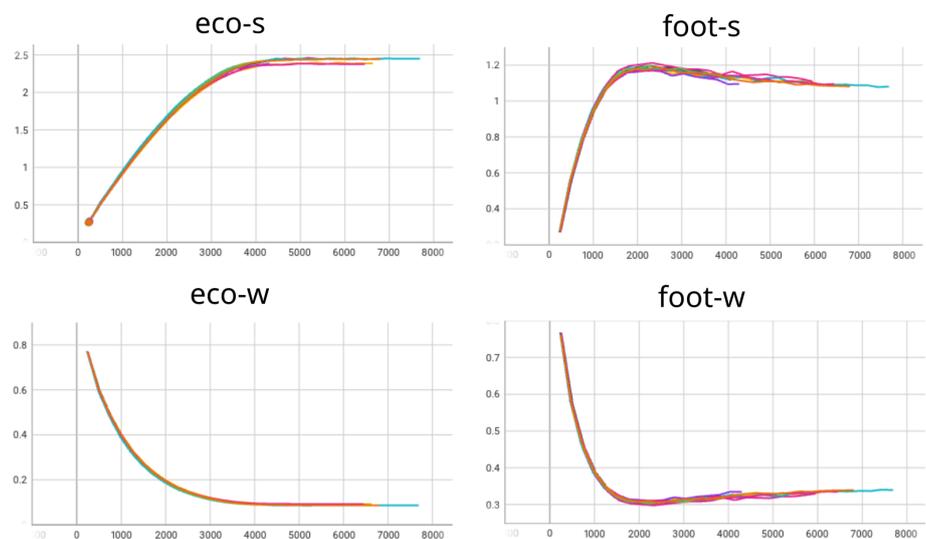**Figure A3.** Eddy-covariance towers used in training.

Figure A4. Evolution of weighting-scheme variables during model training. The colored lines represent the 10 members from the k-folds cross validation.
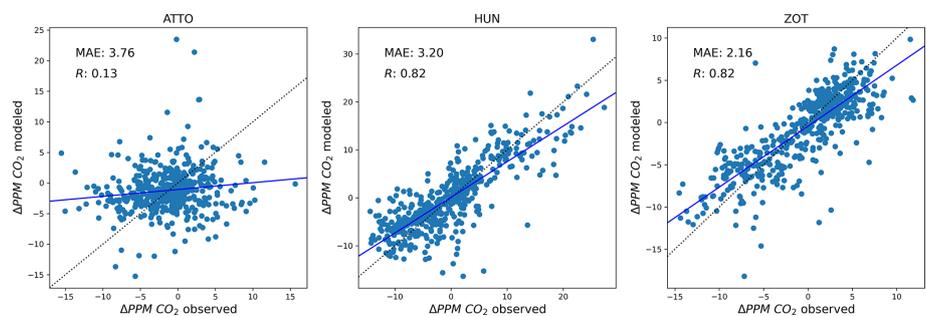

Figure A5. Per-tower representative STILT operator training performance (estimated $\Delta\mathrm{PPM\,CO_2}$ and observed $\Delta\mathrm{PPM\,CO_2}$) for the final epoch of one ensemble member. The black, dotted line is the one-to-one line. The blue line is the regression line. The mean absolute error (MSE) and Pearson's R are reported for each tower.

Appendix B: Results
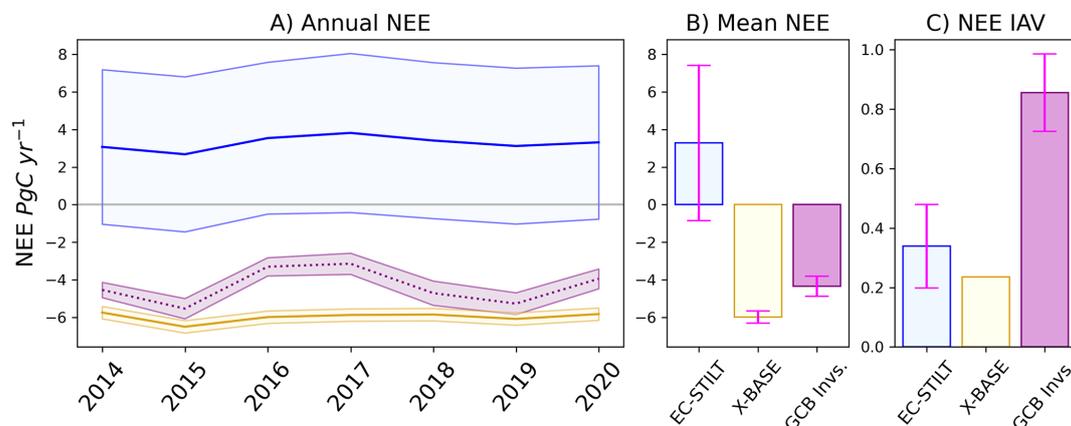
Figure B1. Model ensemble trained without STILT operator: **(A)** Annual NEE integrated globally in PgC yr^{-1}. The blue line is the mean of the EC-NO-STILT ensemble with the shaded area representing the 1σ range across ensemble members. The yellow line is FLUXCOM X-BASE. The purple dotted line is the mean of the GCB23 ensemble of atmospheric inversions, and the shaded area is the standard deviation across the ensemble members. **(B)** Mean NEE integrated globally over all years. The error bars represent the 1σ range across ensemble members or published uncertainty. **(C)** IAV of NEE, calculated as the standard deviation of the global integral. The error bars represent the 1σ range of IAV across ensemble members.
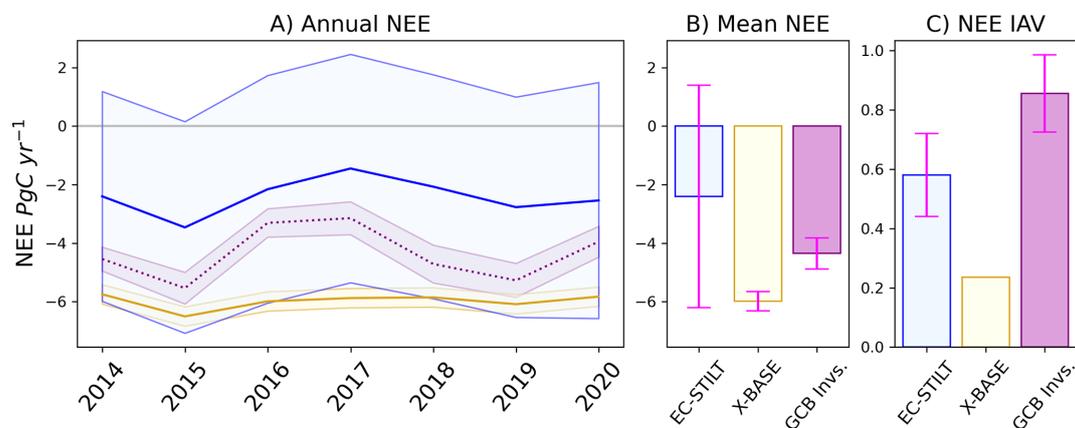
Figure B2. Model ensemble trained a mean LBC term: **(A)** Annual NEE integrated globally in PgC yr^{-1}. The blue line is the mean of the meanLBC ensemble ($n = 3$) with the shaded area representing the 1σ range across ensemble members. The yellow line is FLUXCOM X-BASE. The purple dotted line is the mean of the GCB23 ensemble of atmospheric inversions, and the shaded area is the standard deviation across the ensemble members. **(B)** Mean NEE integrated globally over all years. The error bars represent the 1σ range across ensemble members or published uncertainty. **(C)** IAV of NEE, calculated as the standard deviation of the global integral. The error bars represent the 1σ range of IAV across ensemble members.

Table B1. TA/VPD: Kullback-Leibler distance (KLD) and Jennsson-Shannon metric (JS) between the estimated global and regional PDFs of training data subsets and the estimated PDF of the natural distribution. Columns represent the eddy-covariance (EC) training data, EC plus all three tall towers (EC + Tower), and EC training data plus single towers (e.g. ATTO represents EC plus ATTO), and plus two towers, notated by the tower removed from consideration (e.g. NO-ATTO represents EC + HUN + ZOT). The bold entry represents the lowest distance by metric. Multiple bold entries indicate identical performance across different training sets.

	KLD								JS							
	EC	EC+Tower	ATTO	HUN	ZOT	NO-ATTO	NO-HUN	NO-ZOT	EC	EC+Tower	ATTO	HUN	ZOT	NO-ATTO	NO-HUN	NO-ZOT
Global	0.787	**0.607**	1.181	1.134	1.139	0.879	0.834	0.834	0.407	**0.379**	0.488	0.509	0.514	0.474	0.411	0.437
Af	83.567	47.071	**23.417**	77.017	58.937	67.289	34.966	45.224	0.765	0.733	**0.671**	0.805	0.798	0.810	0.706	0.718
Am	66.614	32.756	**11.305**	76.084	66.614	76.084	**11.305**	32.756	0.731	0.620	**0.553**	0.799	0.731	0.799	**0.553**	0.620
Aw	24.233	12.000	**5.395**	22.279	24.233	22.279	**5.395**	12.000	0.730	0.524	**0.393**	0.791	0.730	0.791	**0.393**	0.524
BWh	5.851	4.714	**3.555**	5.486	9.354	6.783	5.367	3.930	0.731	**0.703**	0.716	0.731	0.764	0.741	0.720	0.697
BWk	5.092	4.958	9.180	**4.119**	5.117	4.184	6.249	5.590	**0.624**	0.694	0.642	0.665	0.737	0.696	0.731	0.666
BSh	4.124	3.963	**2.828**	6.059	4.124	6.059	**2.828**	3.963	0.627	**0.601**	0.611	0.700	0.627	0.700	0.611	**0.601**
BSk	3.375	3.103	4.981	**2.791**	3.554	2.827	3.914	3.359	**0.550**	0.631	0.581	0.604	0.632	0.632	0.676	0.608
Csa	8.749	13.330	33.497	**3.348**	8.671	4.204	22.528	15.308	0.525	0.571	0.639	**0.501**	0.660	0.549	0.674	0.541
Csb	4.470	3.812	7.700	**1.953**	5.098	3.037	5.653	3.678	0.402	0.509	0.543	**0.369**	0.658	0.503	0.640	0.418
Cwa	6.536	4.433	**3.662**	4.069	7.389	5.407	5.291	3.252	0.630	0.608	**0.517**	0.660	0.752	0.703	0.620	0.549
Cwb	3.492	3.380	6.333	**1.807**	4.105	2.673	4.562	3.540	0.451	0.548	0.616	**0.409**	0.680	0.538	0.663	0.478
Cfa	9.785	5.051	7.139	3.910	8.174	5.332	7.411	**3.676**	0.603	0.623	0.557	0.583	0.746	0.651	0.701	**0.550**
Cfb	19.391	11.567	11.170	9.489	22.332	13.629	14.793	**8.845**	0.494	0.607	0.685	**0.486**	0.654	0.554	0.705	0.578
Cfc	21.696	81.414	**21.551**	28.290	99.842	81.567	99.576	28.199	**0.587**	0.720	0.588	0.622	0.759	0.721	0.758	0.622
Dsa	76.648	47.597	82.372	70.603	**23.326**	46.323	26.371	72.161	0.734	0.776	**0.724**	0.779	0.775	0.779	0.769	0.775
Dsb	**2.251**	2.518	2.352	2.877	2.604	2.521	2.578	2.878	0.532	0.621	**0.528**	0.641	0.629	0.624	0.622	0.637
Dsc	9.490	4.868	9.600	9.843	**2.632**	4.833	2.696	9.876	0.680	0.628	0.684	0.721	**0.592**	0.628	0.593	0.722
Dwa	2.521	1.183	2.747	1.405	1.784	**1.140**	1.849	1.462	0.440	**0.410**	0.440	0.469	0.483	0.412	0.473	0.467
Dwb	2.985	2.869	3.090	3.844	2.597	2.865	**2.593**	3.851	0.565	0.588	0.570	0.659	0.567	0.589	**0.565**	0.658
Dwc	15.161	6.694	15.260	16.198	**4.363**	6.665	4.405	16.231	**0.680**	0.706	0.682	0.736	0.694	0.706	0.694	0.736
Dwd	**147.545**	277.802	**147.545**	**147.545**	277.802	277.802	277.802	**147.545**	**0.745**	0.814	**0.745**	**0.745**	0.814	0.814	0.814	**0.745**
Dfa	3.835	1.924	4.163	**1.776**	3.887	1.834	4.146	1.871	0.513	0.429	0.508	0.425	0.541	0.430	0.535	**0.424**
Dfb	3.027	1.999	3.299	2.775	**1.968**	1.975	1.992	2.828	0.474	0.468	0.484	0.539	0.479	**0.468**	0.477	0.540
Dfc	5.369	2.710	5.510	5.860	**1.516**	2.690	1.550	5.890	0.625	0.555	0.631	0.672	**0.511**	0.554	0.512	0.673
Dfd	**68.152**	176.356	**68.152**	**68.152**	176.356	176.356	176.356	**68.152**	**0.716**	0.819	**0.716**	**0.716**	0.819	0.819	0.819	**0.716**

Table B2. EVI/NDWI: Kullback-Leibler distance (KLD) and Jennsson-Shannon metric (JS) between the estimated global and regional PDFs of training data subsets and the estimated PDF of the natural distribution. Columns represent the eddy-covariance (EC) training data, EC plus all three tall towers (EC + Tower), and EC training data plus single towers (e.g. ATTO represents EC plus ATTO), and plus two towers, notated by the tower removed from consideration (e.g. NO-ATTO represents EC + HUN + ZOT). The bold entry represents the lowest distance by metric.

	KLD								JS							
	EC	EC+Tower	ATTO	HUN	ZOT	NO-ATTO	NO-HUN	NO-ZOT	EC	EC+Tower	ATTO	HUN	ZOT	NO-ATTO	NO-HUN	NO-ZOT
Global	0.439	**0.251**	0.767	0.396	0.536	0.348	0.317	0.435	0.323	**0.261**	0.422	0.317	0.366	0.303	0.285	0.333
Af	2.697	2.121	**1.244**	2.761	3.536	3.097	1.969	1.705	0.690	0.580	**0.437**	0.700	0.747	0.724	0.546	0.523
Am	1.759	1.188	**0.608**	1.563	2.554	1.993	1.176	0.797	0.614	0.474	**0.336**	0.592	0.692	0.635	0.459	0.400
Aw	1.344	0.872	0.574	0.806	2.336	1.507	1.079	**0.415**	0.482	0.375	0.318	0.391	0.619	0.498	0.420	**0.275**
BWh	10.913	7.793	11.561	**5.189**	9.092	6.568	9.729	7.935	**0.664**	0.685	0.722	0.672	0.671	0.668	0.692	0.698
BWk	11.947	27.308	59.971	7.925	**6.423**	6.492	35.423	37.891	0.697	0.717	0.757	0.709	**0.691**	0.698	0.720	0.736
BSh	3.052	3.076	3.947	**2.308**	3.875	3.006	3.596	2.962	**0.571**	0.588	0.600	0.572	0.632	0.596	0.600	0.581
BSk	3.339	5.185	10.533	2.762	2.639	**2.502**	6.469	6.848	**0.607**	0.635	0.690	0.622	0.608	0.611	0.641	0.659
Csa	1.520	3.673	8.317	1.085	1.374	**1.071**	4.924	5.071	0.416	0.476	0.608	**0.404**	0.448	0.406	0.516	0.516
Csb	**0.331**	0.680	1.495	0.375	0.824	0.517	0.900	0.880	**0.277**	0.368	0.489	0.306	0.419	0.345	0.407	0.399
Cwa	1.219	0.942	1.148	**0.591**	2.040	1.235	1.269	0.665	0.422	0.385	0.405	0.322	0.574	0.442	0.448	**0.318**
Cwb	0.995	0.883	1.529	**0.542**	1.456	0.895	1.192	0.886	0.366	0.392	0.477	**0.326**	0.488	0.392	0.443	0.387
Cfa	0.937	0.963	1.584	**0.421**	1.689	0.969	1.358	0.867	0.408	0.393	0.447	**0.277**	0.564	0.416	0.469	0.337
Cfb	0.522	0.502	0.837	**0.349**	1.146	0.666	0.690	0.453	0.350	0.342	0.416	**0.289**	0.511	0.389	0.402	0.315
Cfc	1.512	2.783	5.946	1.640	0.964	**1.129**	3.340	4.023	0.524	0.544	0.681	0.537	**0.462**	0.481	0.550	0.617
Dsa	51.071	70.147	125.589	**34.553**	39.909	35.320	86.611	85.742	**0.706**	0.738	0.776	0.727	0.705	0.716	0.739	0.757
Dsb	1.798	5.205	12.020	1.937	**0.833**	1.186	6.691	7.651	0.458	0.521	0.656	0.516	**0.415**	0.452	0.521	0.601
Dsc	1.774	2.297	4.668	1.847	**1.009**	1.242	2.600	3.313	0.510	0.533	0.682	0.535	**0.422**	0.461	0.535	0.618
Dwa	8.745	17.152	37.154	**4.797**	6.165	4.944	22.693	23.012	0.546	0.580	0.645	**0.539**	0.577	0.551	0.604	0.593
Dwb	1.727	3.390	7.812	1.484	**1.081**	1.099	4.356	4.889	0.464	0.495	0.620	0.461	0.446	**0.436**	0.518	0.546
Dwc	3.043	7.168	16.498	3.119	**0.908**	1.721	9.080	10.637	0.516	0.523	0.688	0.541	**0.369**	0.439	0.517	0.625
Dwd	15.213	53.923	121.562	15.996	**4.628**	9.252	69.025	78.036	0.598	0.606	0.734	0.624	**0.493**	0.547	0.598	0.686
Dfa	2.893	2.863	4.060	**1.813**	3.586	2.584	3.542	2.791	0.533	0.550	0.563	0.501	0.627	0.557	0.584	**0.522**
Dfb	0.623	1.581	4.260	0.602	**0.258**	0.273	2.107	2.553	0.306	0.355	0.567	0.313	0.244	**0.226**	0.389	0.461
Dfc	1.646	2.393	5.245	1.999	**0.471**	1.027	2.641	3.745	0.484	0.497	0.684	0.523	**0.310**	0.399	0.487	0.618
Dfd	37.695	90.866	191.052	36.108	**12.521**	21.784	111.177	129.113	0.667	0.683	0.764	0.695	**0.612**	0.647	0.675	0.736

Table B3. EC-only/Tower-enriched: Kullback-Leibler distance (KLD) and Jennsson-Shannon metric (JS) between the estimated global and regional PDFs of training data subsets and the estimated PDF of the natural distribution. For each set of variables and metric, columns represent the eddy-covariance (EC) training data, and the EC plus all three tall towers (EC + Tower). The bold entry represents the lowest distance by metric and variable set.

	TA/VPD				EVI/NDWI			
	KLD		JS		KLD		JS	
	EC	EC + Tower	EC	EC + Tower	EC	EC + Tower	EC	EC + Tower
Global	0.787	**0.607**	0.407	**0.379**	0.439	**0.251**	0.323	**0.261**
Af	83.567	**47.071**	0.765	**0.733**	2.697	**2.121**	0.690	**0.580**
Am	66.614	**32.756**	0.731	**0.620**	1.759	**1.188**	0.614	**0.474**
Aw	24.233	**12.000**	0.730	**0.524**	1.344	**0.872**	0.482	**0.375**
BWh	5.851	**4.714**	0.731	**0.703**	10.913	**7.793**	0.664	0.685
BWk	5.092	**4.958**	**0.624**	0.694	**11.947**	27.308	**0.697**	0.717
BSh	4.124	**3.963**	0.627	**0.601**	**3.052**	3.076	**0.571**	0.588
BSk	3.375	**3.103**	**0.550**	0.631	**3.339**	5.185	**0.607**	0.635
Csa	**8.749**	13.330	**0.525**	0.571	**1.520**	3.673	**0.416**	0.476
Csb	4.470	**3.812**	0.402	0.509	**0.331**	0.680	**0.277**	0.368
Cwa	6.536	**4.433**	0.630	**0.608**	1.219	**0.942**	**0.422**	0.385
Cwb	3.492	**3.380**	**0.451**	0.548	0.995	**0.883**	**0.366**	0.392
Cfa	9.785	**5.051**	**0.603**	0.623	**0.937**	0.963	0.408	**0.393**
Cfb	19.391	**11.567**	**0.494**	0.607	0.522	**0.502**	0.350	**0.342**
Cfc	**21.696**	81.414	**0.587**	0.720	**1.512**	2.783	**0.524**	0.544
Dsa	76.648	**47.597**	**0.734**	0.776	**51.071**	70.147	**0.706**	0.738
Dsb	**2.251**	2.518	**0.532**	0.621	**1.798**	5.205	**0.458**	0.521
Dsc	9.490	**4.868**	0.680	**0.628**	**1.774**	2.297	**0.510**	0.533
Dwa	2.521	**1.183**	0.440	**0.410**	**8.745**	17.152	**0.546**	0.580
Dwb	2.985	**2.869**	**0.565**	0.588	**1.727**	3.390	**0.464**	0.495
Dwc	15.161	**6.694**	**0.680**	0.706	**3.043**	7.168	**0.516**	0.523
Dwd	**147.545**	277.802	**0.745**	0.814	**15.213**	53.923	**0.598**	0.606
Dfa	3.835	**1.924**	0.513	**0.429**	2.893	**2.863**	**0.533**	0.550
Dfb	3.027	**1.999**	0.474	**0.468**	**0.623**	1.581	**0.306**	0.355
Dfc	5.369	**2.710**	0.625	**0.555**	**1.646**	2.393	**0.484**	0.497
Dfd	**68.152**	176.356	**0.716**	0.819	**37.695**	90.866	**0.667**	0.683

Table B4. Cross-validation results for EC-STILT and X-BASE. The columns are hourly and daily R^2 and NSE for the sites that were reserved for validation during testing.

	Hourly NEE		Daily NEE	
	R^2	NSE	R^2	NSE
EC-STILT	0.77	0.66	0.71	0.71
X-BASE	0.71	0.71	0.81	0.81

References

Agarap, A. F.: Deep Learning using Rectified Linear Units (ReLU), arXiv:1803.08375 [cs, stat], https://doi.org/10.48550/arXiv.1803.08375, 2019.

Ahlström, A., Raupach, M. R., Schurgers, G., Smith, B., Arneth, A., Jung, M., Reichstein, M., Canadell, J. G., Friedlingstein, P., Jain, A. K., Kato, E., Poulter, B., Sitch, S., Stocker, B. D., Viovy, N., Wang, Y. P., Wiltshire, A., Zaehle, S., and Zeng, N.: The dominant role of semi-arid ecosystems in the trend and variability of the land CO_2 sink, Science, 348, 895–899, https://doi.org/10.1126/science.aaa1668, 2015.

Amiro, B.: FLUXNET2015 CA-SF2 Saskatchewan – Western Boreal, Forest Burned in 1989, FLUXNET [data set], https://doi.org/10.18140/flx/1440047 (last access: 1 November 2022), 2016a.

Amiro, B.: FLUXNET2015 CA-Man Manitoba – Northern Old Black Spruce (Former BOREAS Northern Study Area), FLUXNET [data set], https://doi.org/10.18140/flx/1440035 (last access: 1 November 2022), 2016b.

Amiro, B.: FLUXNET2015 CA-SF1 Saskatchewan – Western Boreal, Forest Burned in 1977, FLUXNET [data set], https://doi.org/10.18140/flx/1440046 (last access: 1 November 2022), 2016c.

Amiro, B.: FLUXNET2015 CA-SF3 Saskatchewan – Western Boreal, Forest Burned in 1998, FLUXNET [data set], https://doi.org/10.18140/flx/1440048 (last access: 1 November 2022), 2016d.

Ammann, C.: FLUXNET2015 CH-Oe1 Oensingen Grassland, FLUXNET [data set], https://doi.org/10.18140/flx/1440135 (last access: 1 November 2022), 2016.

Angevine, W. M., Brioude, J., McKeen, S., and Holloway, J. S.: Uncertainty in Lagrangian pollutant transport simulations due to meteorological uncertainty from a mesoscale WRF ensemble, Geosci. Model Dev., 7, 2817–2829, https://doi.org/10.5194/gmd-7-2817-2014, 2014.

Arain, M.: AmeriFlux FLUXNET-1F CA-TPD Ontario – Turkey Point Mature Deciduous, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881567 (last access: 1 November 2022), 2022a.

Arain, M.: AmeriFlux FLUXNET-1F CA-TP3 Ontario – Turkey Point 1974 Plantation White Pine, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881566 (last access: 1 November 2022), 2022b.

Arain, M. A.: FLUXNET2015 CA-TP2 Ontario – Turkey Point 1989 Plantation White Pine, FLUXNET [data set], https://doi.org/10.18140/flx/1440051 (last access: 1 November 2022), 2016a.

Arain, M. A.: FLUXNET2015 CA-TP1 Ontario – Turkey Point 2002 Plantation White Pine, FLUXNET [data set], https://doi.org/10.18140/flx/1440050 (last access: 1 November 2022), 2016b.

Arain, M. A.: FLUXNET2015 CA-TP4 Ontario – Turkey Point 1939 Plantation White Pine, FLUXNET [data set], https://doi.org/10.18140/flx/1440053 (last access: 1 November 2022), 2016c.

Ardö, J., El Tahir, B. A., and ElKhidir, H. A. M.: FLUXNET2015 SD-Dem Demokeya, FLUXNET [data set],

https://doi.org/10.18140/flx/1440186 (last access: 1 November 2022), 2016.

Arndt, S., Hinko-Najera, N., and Griebel, A.: FLUXNET2015 AU-Wom Wombat, FLUXNET [data set], https://doi.org/10.18140/flx/1440207 (last access: 1 November 2022), 2016.

Aubinet, M., Berbigier, P., Bernhofer, C., Cescatti, A., Feigenwinter, C., Granier, A., Grünwald, T., Havrankova, K., Heinesch, B., Longdoz, B., Marcolla, B., Montagnani, L., and Sedlak, P.: Comparing CO2 Storage and Advection Conditions at Night at Different Carboeuroflux Sites, Bound.-Lay. Meteorol., 116, 63–93, https://doi.org/10.1007/s10546-004-7091-8, 2005.

Aurela, M., Lohila, A., Tuovinen, J.-P., Hatakka, J., Rainne, J., Mäkelä, T., and Lauria, T.: FLUXNET2015 FI-Lom Lompolojankka, FLUXNET [data set], https://doi.org/10.18140/flx/1440228 (last access: 1 November 2022), 2016a.

Aurela, M., Tuovinen, J.-P., Hatakka, J., Lohila, A., Mäkelä, T., Rainne, J., and Lauria, T.: FLUXNET2015 FI-Sod Sodankyla, FLUXNET [data set], https://doi.org/10.18140/flx/1440160 (last access: 1 November 2022), 2016b.

Badgley, G., Anderegg, L. D. L., Berry, J. A., and Field, C. B.: Terrestrial gross primary production: Using NIRV to scale from site to globe, Global Change Biol., 25, 3731–3740, https://doi.org/10.1111/gcb.14729, 2019.

Baker, D. F., Law, R. M., Gurney, K. R., Rayner, P., Peylin, P., Denning, A. S., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyutov, S., Masarie, K., Prather, M., Pak, B., Taguchi, S., and Zhu, Z.: TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO2 fluxes, 1988–2003, Global Biogeochem. Cycles, 20, https://doi.org/10.1029/2004GB002439, 2006.

Baker, J. and Griffis, T.: AmeriFlux FLUXNET-1F US-Ro5 Rosemount I18South, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1818371 (last access: 1 November 2022), 2021.

Baker, J. and Griffis, T.: AmeriFlux FLUXNET-1F US-Ro4 Rosemount Prairie, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881589 (last access: 1 November 2022), 2022a.

Baker, J. and Griffis, T.: AmeriFlux FLUXNET-1F US-Ro6 Rosemount I18North, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881590 (last access: 1 November 2022), 2022b.

Baker, J., Griffis, T., and Griffis, T.: AmeriFlux FLUXNET-1F US-Ro1 Rosemount- G21, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881588 (last access: 1 November 2022), 2022.

Baldocchi, D.: FLUXNET2015 US-Twt Twitchell Island, FLUXNET [data set], https://doi.org/10.18140/flx/1440106 (last access: 1 November 2022), 2016.

Baldocchi, D. and Ma, S.: FLUXNET2015 US-Ton Tonzi Ranch, FLUXNET [data set], https://doi.org/10.18140/flx/1440092 (last access: 1 November 2022), 2016.

Baldocchi, D., Ma, S., and Xu, L.: FLUXNET2015 US-Var Vaira Ranch- Ione, FLUXNET [data set], https://doi.org/10.18140/flx/1440094 (last access: 1 November 2022), 2016.

Bastos, A., Ciais, P., Sitch, S., Aragão, L. E. O. C., Chevallier, F., Fawcett, D., Rosan, T. M., Saunois, M., Günther, D., Perugini, L., Robert, C., Deng, Z., Pongratz, J., Ganzenmüller, R., Fuchs, R., Winkler, K., Zaehle, S., and Albergel, C.: On the use of Earth Observation to support estimates of national greenhouse gas emissions and sinks for the Global stocktake process: lessons learned from ESA-CCI RECCAP2, Carbon Balance and Management, 17, 15, https://doi.org/10.1186/s13021-022-00214-w, 2022.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F.: Present and future Köppen-Geiger climate classification maps at 1-km resolution, Sci. Data, 5, 180214, https://doi.org/10.1038/sdata.2018.214, 2018.

Belelli, L., Papale, D., and Valentini, R.: FLUXNET2015 RU-Ha1 Hakasia Steppe, FLUXNET [data set], https://doi.org/10.18140/flx/1440184 (last access: 1 November 2022), 2016.

Berbigier, P. and Loustau, D.: FLUXNET2015 FR-LBr Le Bray, FLUXNET [data set], https://doi.org/10.18140/flx/1440163 (last access: 1 November 2022), 2016.

Beringer, J. and Hutley, L.: FLUXNET2015 AU-Fog Fogg Dam, FLUXNET [data set], https://doi.org/10.18140/flx/1440124 (last access: 1 November 2022), 2016a.

Beringer, J. and Hutley, L.: FLUXNET2015 AU-DaP Daly River Savanna, FLUXNET [data set], https://doi.org/10.18140/flx/1440123 (last access: 1 November 2022), 2016b.

Beringer, J. and Hutley, L.: FLUXNET2015 AU-Ade Adelaide River, FLUXNET [data set], https://doi.org/10.18140/flx/1440193 (last access: 1 November 2022), 2016c.

Beringer, J. and Hutley, L.: FLUXNET2015 AU-RDF Red Dirt Melon Farm, Northern Territory, FLUXNET [data set], https://doi.org/10.18140/flx/1440201 (last access: 1 November 2022), 2016d.

Beringer, J. and Hutley, L.: FLUXNET2015 AU-Dry Dry River, FLUXNET [data set], https://doi.org/10.18140/flx/1440197 (last access: 1 November 2022), 2016e.

Beringer, J. and Hutley, P. L.: FLUXNET2015 AU-DaS Daly River Cleared, FLUXNET [data set], https://doi.org/10.18140/flx/1440122 (last access: 1 November 2022), 2016f.

Beringer, J. and Walker, J.: FLUXNET2015 AU-Ync Jaxa, FLUXNET [data set], https://doi.org/10.18140/flx/1440208 (last access: 1 November 2022), 2016.

Beringer, J., Cunningham, S., Baker, P., Cavagnaro, T., MacNally, R., Thompson, R., and McHugh, I.: FLUXNET2015 AU-Whr Whroo, FLUXNET [data set], https://doi.org/10.18140/flx/1440206 (last access: 1 November 2022), 2016a.

Beringer, J., Hutley, L., McGuire, D., and U, P.: FLUXNET2015 AU-Wac Wallaby Creek, FLUXNET [data set], https://doi.org/10.18140/flx/1440127 (last access: 1 November 2022), 2016b.

Bernhofer, C., Grünwald, T., Moderow, U., Hehn, M., Eichelmann, U., and Prasse, H.: FLUXNET2015 DE-Spw Spreewald, FLUXNET [data set], https://doi.org/10.18140/flx/1440220 (last access: 1 November 2022), 2016.

Billesbach, D., Bradford, J., and Torn, M.: FLUXNET2015 US-AR2 ARM USDA UNL OSU Woodward Switchgrass 2,

FLUXNET [data set], https://doi.org/10.18140/flx/1440104 (last access: 1 November 2022), 2016a.

Billesbach, D., Bradford, J., and Torn, M.: FLUXNET2015 US-AR1 ARM USDA UNL OSU Woodward Switchgrass 1, FLUXNET [data set], https://doi.org/10.18140/flx/1440103 (last access: 1 November 2022), 2016b.

Billesbach, D., Kueppers, L., Torn, M., and Biraud, S.: AmeriFlux FLUXNET-1F US-A32 ARM-SGP Medford Hay Pasture, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881568 (last access: 1 November 2022), 2022.

Biraud, S., Fischer, M., Chan, S., and Torn, M.: AmeriFlux FLUXNET-1F US-ARM ARM Southern Great Plains Site – Lamont, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1854366 (last access: 1 November 2022), 2022.

Black, T.: AmeriFlux FLUXNET-1F CA-LP1 British Columbia – Mountain Pine Beetle-Attacked Lodgepole Pine Stand, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1832155 (last access: 1 November 2022), 2021.

Black, T. A.: FLUXNET2015 CA-Obs Saskatchewan – Western Boreal, Mature Black Spruce, FLUXNET [data set], https://doi.org/10.18140/flx/1440044 (last access: 1 November 2022), 2016a.

Black, T. A.: FLUXNET2015 CA-Oas Saskatchewan – Western Boreal, Mature Aspen, FLUXNET [data set], https://doi.org/10.18140/flx/1440043 (last access: 1 November 2022), 2016b.

Blanken, P., Monson, R., Burns, S., Bowling, D., and Turnipseed, A.: AmeriFlux FLUXNET-1F US-NR1 Niwot Ridge Forest (LTER NWT1), AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1871141 (last access: 1 November 2022), 2022.

Bodesheim, P., Jung, M., Gans, F., Mahecha, M. D., and Reichstein, M.: Upscaled diurnal cycles of land–atmosphere fluxes: a new global half-hourly data product, Earth Syst. Sci. Data, 10, 1327–1365, https://doi.org/10.5194/essd-10-1327-2018, 2018.

Bohrer, G.: AmeriFlux FLUXNET-1F US-ORv Olentangy River Wetland Research Park, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1832164 (last access: 1 November 2022), 2021.

Bohrer, G.: AmeriFlux FLUXNET-1F US-UM3 Douglas Lake, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881596 (last access: 1 November 2022), 2022.

Bohrer, G. and Kerns, J.: AmeriFlux FLUXNET-1F US-OWC Old Woman Creek, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1871142 (last access: 1 November 2022), 2022.

Boike, J., Westermann, S., Lüers, J., Langer, M., and Piel, K.: FLUXNET2015 SJ-Blv Bayelva, Spitsbergen, FLUXNET [data set], https://doi.org/10.18140/flx/1440242 (last access: 1 November 2022), 2016.

Botia, S., Komiya, S., Marshall, J., Koch, T., Galkowski, M., Lavric, J., Gomes-Alves, E., Walter, D., Fisch, G., Pinho, D. M., Nelson, B. W., Martins, G., Luijkx, I. T., Koren, G., Florentie, L., de Araujo, A. C., Sa, M., Andreae, M. O., Heimann, M., Peters, W., and Gerbig, C.: The CO2 record at the Amazon Tall Tower Observatory: A new opportunity to study processes on seasonal and inter-annual scales, Global Change Biol., 28, 588–611, https://doi.org/10.1111/gcb.15905, 2022.

Botía, S., Munassar, S., Koch, T., Custodio, D., Basso, L. S., Komiya, S., Lavric, J. V., Walter, D., Gloor, M., Martins, G., Naus, S., Koren, G., Luijkx, I. T., Hantson, S., Miller, J. B., Peters, W., Rödenbeck, C., and Gerbig, C.: Combined CO_2 measurement record indicates Amazon forest carbon uptake is offset by savanna carbon release, Atmos. Chem. Phys., 25, 6219–6255, https://doi.org/10.5194/acp-25-6219-2025, 2025.

Bowling, D.: FLUXNET2015 US-Cop Corral Pocket, FLUXNET [data set], https://doi.org/10.18140/flx/1440100 (last access: 1 November 2022), 2016.

Brunsell, N.: AmeriFlux FLUXNET-1F US-KFS Kansas Field Station, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881585 (last access: 1 November 2022), 2022a.

Brunsell, N.: AmeriFlux FLUXNET-1F US-KLS Kansas Land Institute, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1854367 (last access: 1 November 2022), 2022b.

Cescatti, A., Marcolla, B., Zorer, R., and Gianelle, D.: FLUXNET2015 IT-La2 Lavarone2, FLUXNET [data set], https://doi.org/10.18140/flx/1440235 (last access: 1 November 2022), 2016.

Chamberlain, S., Oikawa, P., Sturtevant, C., Szutu, D., Verfaillie, J., and Baldocchi, D.: AmeriFlux FLUXNET-1F US-Tw3 Twitchell Alfalfa, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881594 (last access: 1 November 2022), 2022.

Chen, J.: FLUXNET2015 US-Wi5 Mixed Young Jack Pine (MYJP), FLUXNET [data set], https://doi.org/10.18140/flx/1440059 (last access: 1 November 2022), 2016a.

Chen, J.: FLUXNET2015 US-Wi3 Mature Hardwood (MHW), FLUXNET [data set], https://doi.org/10.18140/flx/1440057 (last access: 1 November 2022), 2016b.

Chen, J.: FLUXNET2015 US-Wi8 Young Hardwood Clearcut (YHW), FLUXNET [data set], https://doi.org/10.18140/flx/1440062 (last access: 1 November 2022), 2016c.

Chen, J.: FLUXNET2015 US-Wi4 Mature Red Pine (MRP), FLUXNET [data set], https://doi.org/10.18140/flx/1440058 (last access: 1 November 2022), 2016d.

Chen, J.: FLUXNET2015 US-Wi1 Intermediate Hardwood (IHW), FLUXNET [data set], https://doi.org/10.18140/flx/1440054 (last access: 1 November 2022), 2016e.

Chen, J.: FLUXNET2015 US-Wi9 Young Jack Pine (YJP), FLUXNET [data set], https://doi.org/10.18140/flx/1440063 (last access: 1 November 2022), 2016f.

Chen, J.: FLUXNET2015 US-Wi0 Young Red Pine (YRP), FLUXNET [data set], https://doi.org/10.18140/flx/1440055 (last access: 1 November 2022), 2016g.

Chen, J.: FLUXNET2015 US-Wi6 Pine Barrens 1 (PB1), FLUXNET [data set], https://doi.org/10.18140/flx/1440060 (last access: 1 November 2022), 2016h.

Chen, J.: FLUXNET2015 US-Wi7 Red Pine Clearcut (RPCC), FLUXNET [data set], https://doi.org/10.18140/flx/1440061 (last access: 1 November 2022), 2016i.

Chen, J.: FLUXNET2015 US-Wi2 Intermediate Red Pine (IRP), FLUXNET [data set], https://doi.org/10.18140/flx/1440056 (last access: 1 November 2022), 2016j.

Chen, J. and Chu, H.: FLUXNET2015 US-WPT Winous Point North Marsh, FLUXNET [data set], https://doi.org/10.18140/flx/1440116 (last access: 1 November 2022), 2016a.

Chen, J. and Chu, H.: FLUXNET2015 US-CRT Curtice Walter-Berger Cropland, FLUXNET [data set], https://doi.org/10.18140/flx/1440117 (last access: 1 November 2022), 2016b.

Chen, J., Chu, H., and Noormets, A.: FLUXNET2015 US-Oho Oak Openings, FLUXNET [data set], https://doi.org/10.18140/flx/1440088 (last access: 1 November 2022), 2016.

Chen, S.: FLUXNET2015 CN-Du2 Duolungrassland (D01), FLUXNET [data set], https://doi.org/10.18140/flx/1440140 (last access: 1 November 2022), 2016k.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, pp. 785–794, Association for Computing Machinery, New York, NY, USA, ISBN 978-1-4503-4232-2, https://doi.org/10.1145/2939672.2939785, 2016.

Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F.-M., Chédin, A., and Ciais, P.: Inferring CO2 sources and sinks from satellite observations: Method and application to TOVS data, J. Geophys. Res.-Atmos., 110, https://doi.org/10.1029/2005JD006390, 2005.

Chevallier, F., Remaud, M., O'Dell, C. W., Baker, D., Peylin, P., and Cozic, A.: Objective evaluation of surface- and satellite-driven carbon dioxide atmospheric inversions, Atmos. Chem. Phys., 19, 14233–14251, https://doi.org/10.5194/acp-19-14233-2019, 2019.

Christen, A. and Knox, S.: AmeriFlux FLUXNET-1F CA-DBB Delta Burns Bog, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881565 (last access: 1 November 2022), 2022.

Christensen, T.: FLUXNET2015 SJ-Adv Adventdalen, FLUXNET [data set], https://doi.org/10.18140/flx/1440241 (last access: 1 November 2022), 2016.

Chu, H., Baldocchi, D. D., John, R., Wolf, S., and Reichstein, M.: Fluxes all of the time? A primer on the temporal representativeness of FLUXNET, J. Geophys. Res.-Biogeo., 122, 289–307, https://doi.org/10.1002/2016JG003576, 2017.

Ciais, P., Bastos, A., Chevallier, F., Lauerwald, R., Poulter, B., Canadell, J. G., Hugelius, G., Jackson, R. B., Jain, A., Jones, M., Kondo, M., Luijkx, I. T., Patra, P. K., Peters, W., Pongratz, J., Petrescu, A. M. R., Piao, S., Qiu, C., Von Randow, C., Regnier, P., Saunois, M., Scholes, R., Shvidenko, A., Tian, H., Yang, H., Wang, X., and Zheng, B.: Definitions and methods to estimate regional land carbon fluxes for the second phase of the REgional Carbon Cycle Assessment and Processes Project (RECCAP-2), Geosci. Model Dev., 15, 1289–1316, https://doi.org/10.5194/gmd-15-1289-2022, 2022.

Cleverly, J. and Eamus, D.: FLUXNET2015 AU-TTE Ti Tree East, FLUXNET [data set], https://doi.org/10.18140/flx/1440205 (last access: 1 November 2022), 2016a.

Cleverly, J. and Eamus, D.: FLUXNET2015 AU-ASM Alice Springs, FLUXNET [data set], https://doi.org/10.18140/flx/1440194 (last access: 1 November 2022), 2016b.

Cochrane, M. A. and Laurance, W. F.: Synergisms among Fire, Land Use, and Climate Change in the Amazon, Ambio, 37, 522–527, 2008.

Crisp, D., Dolman, H., Tanhua, T., McKinley, G. A., Hauck, J., Bastos, A., Sitch, S., Eggleston, S., and Aich, V.: How Well Do We Understand the Land-Ocean-Atmosphere Carbon Cycle?, Rev. Geophys., 60, e2021RG000736, https://doi.org/10.1029/2021RG000736, 2022.

Desai, A.: FLUXNET2015 US-WCr Willow Creek, FLUXNET [data set], https://doi.org/10.18140/flx/1440095 (last access: 1 November 2022), 2016a.

Desai, A.: FLUXNET2015 US-Syv Sylvania Wilderness Area, FLUXNET [data set], https://doi.org/10.18140/flx/1440091 (last access: 1 November 2022), 2016b.

Desai, A.: FLUXNET2015 US-Los Lost Creek, FLUXNET [data set], https://doi.org/10.18140/flx/1440076 (last access: 1 November 2022), 2016c.

Desai, A.: FLUXNET2015 US-PFa Park Falls/WLEF, FLUXNET [data set], https://doi.org/10.18140/flx/1440089 (last access: 1 November 2022), 2016d.

Desai, A.: AmeriFlux FLUXNET-1F US-CS1 Central Sands Irrigated Agricultural Field, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881576 (last access: 1 November 2022), 2022a.

Desai, A.: AmeriFlux FLUXNET-1F US-CS4 Central Sands Irrigated Agricultural Field, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881579 (last access: 1 November 2022), 2022b.

Desai, A.: AmeriFlux FLUXNET-1F US-CS2 Tri County School Pine Forest, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881577 (last access: 1 November 2022), 2022c.

Desai, A.: AmeriFlux FLUXNET-1F US-CS3 Central Sands Irrigated Agricultural Field, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881578 (last access: 1 November 2022), 2022d.

Dolman, H., Hendriks, D., Parmentier, F.-J., Marchesini, L. B., Dean, J., and Van Huissteden, K.: FLUXNET2015 NL-Hor Horstermeer, FLUXNET [data set], https://doi.org/10.18140/flx/1440177 (last access: 1 November 2022), 2016a.

Dolman, H., Van Der Molen, M., Parmentier, F.-J., Marchesini, L. B., Dean, J., Van Huissteden, K., and Maximov, T.: FLUXNET2015 RU-Cok Chokurdakh, FLUXNET [data set], https://doi.org/10.18140/flx/1440182 (last access: 1 November 2022), 2016b.

Dong, G.: FLUXNET2015 CN-Cng Changling, FLUXNET [data set], https://doi.org/10.18140/flx/1440209 (last access: 1 November 2022), 2016.

Drake, B. and Hinkle, R.: FLUXNET2015 US-KS1 Kennedy Space Center (Slash Pine), FLUXNET [data set], https://doi.org/10.18140/flx/1440074 (last access: 1 November 2022), 2016a.

Drake, B. and Hinkle, R.: FLUXNET2015 US-KS2 Kennedy Space Center (Scrub Oak), FLUXNET [data set],

https://doi.org/10.18140/flx/1440075 (last access: 1 November 2022), 2016b.

Euskirchen, E.: AmeriFlux FLUXNET-1F US-BZo Bonanza Creek Old Thermokarst Bog, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881571 (last access: 1 November 2022), 2022a.

Euskirchen, E.: AmeriFlux FLUXNET-1F US-BZB Bonanza Creek Thermokarst Bog, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881569 (last access: 1 November 2022), 2022b.

Euskirchen, E.: AmeriFlux FLUXNET-1F US-BZF Bonanza Creek Rich Fen, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881570 (last access: 1 November 2022), 2022c.

Euskirchen, E.: AmeriFlux FLUXNET-1F US-BZS Bonanza Creek Black Spruce, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881572 (last access: 1 November 2022), 2022d.

Euskirchen, E., Shaver, G., and Bret-Harte, S.: AmeriFlux FLUXNET-1F US-ICs Imnavait Creek Watershed Wet Sedge Tundra, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1871138 (last access: 1 November 2022), 2022a.

Euskirchen, E., Shaver, G., and Bret-Harte, S.: AmeriFlux FLUXNET-1F US-ICt Imnavait Creek Watershed Tussock Tundra, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881583 (last access: 1 November 2022), 2022b.

Ewers, B. and Pendall, E.: FLUXNET2015 US-Sta Saratoga, FLUXNET [data set], https://doi.org/10.18140/flx/1440115 (last access: 1 November 2022), 2016.

Faassen, K. A. P., Vilà-Guerau de Arellano, J., González-Armas, R., Heusinkveld, B. G., Mammarella, I., Peters, W., and Luijkx, I. T.: Separating above-canopy CO_2 and O_2 measurements into their atmospheric and biospheric signatures, Biogeosciences, 21, 3015–3039, https://doi.org/10.5194/bg-21-3015-2024, 2024.

Fares, S.: FLUXNET2015 US-Lin Lindcove Orange Orchard, FLUXNET [data set], https://doi.org/10.18140/flx/1440107 (last access: 1 November 2022), 2016.

Flerchinger, G.: AmeriFlux FLUXNET-1F US-Rwf RCEW Upper Sheep Prescibed Fire, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881591 (last access: 1 November 2022), 2022a.

Flerchinger, G.: AmeriFlux FLUXNET-1F US-Rws Reynolds Creek Wyoming Big Sagebrush, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881592 (last access: 1 November 2022), 2022b.

Flerchinger, G.: AmeriFlux FLUXNET-1F US-Rms RCEW Mountain Big Sagebrush, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881587 (last access: 1 November 2022), 2022c.

Flerchinger, G. and Reba, M.: AmeriFlux FLUXNET-1F US-Rwe RCEW Reynolds Mountain East, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1871143 (last access: 1 November 2022), 2022.

Forsythe, J., Kline, M., and O'Halloran, T.: AmeriFlux FLUXNET-1F US-HB1 North Inlet Crab Haul Creek, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1832160 (last access: 1 November 2022), 2021.

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., and Knutti, R.: Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks, J. Climate, 27, 511–526, https://doi.org/10.1175/JCLI-D-12-00579.1, 2014.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., Landschützer, P., Le Quéré, C., Luijkx, I. T., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., Barbero, L., Bates, N. R., Becker, M., Bellouin, N., Decharme, B., Bopp, L., Brasika, I. B. M., Cadule, P., Chamberlain, M. A., Chandra, N., Chau, T.-T.-T., Chevallier, F., Chini, L. P., Cronin, M., Dou, X., Enyo, K., Evans, W., Falk, S., Feely, R. A., Feng, L., Ford, D. J., Gasser, T., Ghattas, J., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, Ö., Harris, I., Hefner, M., Heinke, J., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Jacobson, A. R., Jain, A., Jarníková, T., Jersild, A., Jiang, F., Jin, Z., Joos, F., Kato, E., Keeling, R. F., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Körtzinger, A., Lan, X., Lefèvre, N., Li, H., Liu, J., Liu, Z., Ma, L., Marland, G., Mayot, N., McGuire, P. C., McKinley, G. A., Meyer, G., Morgan, E. J., Munro, D. R., Nakaoka, S.-I., Niwa, Y., O'Brien, K. M., Olsen, A., Omar, A. M., Ono, T., Paulsen, M., Pierrot, D., Pocock, K., Poulter, B., Powis, C. M., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Rosan, T. M., Schwinger, J., Séférian, R., Smallman, T. L., Smith, S. M., Sospedra-Alfonso, R., Sun, Q., Sutton, A. J., Sweeney, C., Takao, S., Tans, P. P., Tian, H., Tilbrook, B., Tsujino, H., Tubiello, F., van der Werf, G. R., van Ooijen, E., Wanninkhof, R., Watanabe, M., Wimart-Rousseau, C., Yang, D., Yang, X., Yuan, W., Yue, X., Zaehle, S., Zeng, J., and Zheng, B.: Global Carbon Budget 2023, Earth Syst. Sci. Data, 15, 5301–5369, https://doi.org/10.5194/essd-15-5301-2023, 2023.

Fu, Z., Gerken, T., Bromley, G., Araújo, A., Bonal, D., Burban, B., Ficklin, D., Fuentes, J. D., Goulden, M., Hirano, T., Kosugi, Y., Liddell, M., Nicolini, G., Niu, S., Roupsard, O., Stefani, P., Mi, C., Tofte, Z., Xiao, J., Valentini, R., Wolf, S., and Stoy, P. C.: The surface-atmosphere exchange of carbon dioxide in tropical rainforests: Sensitivity to environmental drivers and flux measurement methodology, Agric. Forest Meteorol., 263, 292–307, https://doi.org/10.1016/j.agrformet.2018.09.001, 2018.

Gao, B.-C.: NDWI – A normalized difference water index for remote sensing of vegetation liquid water from space, Remote Sens. Environ., 58, 257–266, https://doi.org/10.1016/S0034-4257(96)00067-3, 1996.

Garcia, A., Di Bella, C., Houspanossian, J., Magliano, P., Jobbágy, E., Posse, G., Fernández, R., and Nosetto, M.: FLUXNET2015 AR-SLu San Luis, FLUXNET [data set], https://doi.org/10.18140/flx/1440191 (last access: 1 November 2022), 2016.

Gaubert, B., Edwards, D. P., Anderson, J. L., Arellano, A. F., Barré, J., Buchholz, R. R., Darras, S., Emmons, L. K., Fillmore, D., Granier, C., Hannigan, J. W., Ortega, I., Raeder, K., Soulié, A., Tang, W., Worden, H. M., and Ziskin, D.: Global Scale Inversions from MOPITT CO and MODIS AOD, Remote Sens., 15, 4813, https://doi.org/10.3390/rs15194813, 2023.

Goldstein, A.: FLUXNET2015 US-Blo Blodgett Forest, FLUXNET [data set], https://doi.org/10.18140/flx/1440068 (last access: 1 November 2022), 2016.

Goslee, S.: AmeriFlux FLUXNET-1F US-HWB USDA ARS Pasture Sytems and Watershed Management Research Unit- Hawbecker Site, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881582 (last access: 1 November 2022), 2022.

Gough, C., Bohrer, G., and Curtis, P.: FLUXNET2015 US-UMB Univ. Of Mich. Biological Station, FLUXNET [data set], https://doi.org/10.18140/flx/1440093 (last access: 1 November 2022), 2016.

Gough, C., Bohrer, G., and Curtis, P.: AmeriFlux FLUXNET-1F US-UMd UMBS Disturbance, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881597 (last access: 1 November 2022), 2022.

Goulden, M.: FLUXNET2015 CA-NS2 UCI-1930 Burn Site, FLUXNET [data set], https://doi.org/10.18140/flx/1440037 (last access: 1 November 2022), 2016a.

Goulden, M.: FLUXNET2015 CA-NS3 UCI-1964 Burn Site, FLUXNET [data set], https://doi.org/10.18140/flx/1440038 (last access: 1 November 2022), 2016b.

Goulden, M.: FLUXNET2015 CA-NS4 UCI-1964 Burn Site Wet, FLUXNET [data set], https://doi.org/10.18140/flx/1440039 (last access: 1 November 2022), 2016c.

Goulden, M.: FLUXNET2015 BR-Sa3 Santarem-Km83-Logged Forest, FLUXNET [data set], https://doi.org/10.18140/flx/1440033 (last access: 1 November 2022), 2016d.

Goulden, M.: FLUXNET2015 CA-NS6 UCI-1989 Burn Site, FLUXNET [data set], https://doi.org/10.18140/flx/1440041 (last access: 1 November 2022), 2016e.

Goulden, M.: FLUXNET2015 CA-NS7 UCI-1998 Burn Site, FLUXNET [data set], https://doi.org/10.18140/flx/1440042 (last access: 1 November 2022), 2016f.

Goulden, M.: FLUXNET2015 CA-NS5 UCI-1981 Burn Site, FLUXNET [data set], https://doi.org/10.18140/flx/1440040 (last access: 1 November 2022), 2016g.

Griffis, T. and Roman, T.: AmeriFlux FLUXNET-1F PE-QFR Quistococha Forest Reserve, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1832157 (last access: 1 November 2022), 2021.

Gruening, C., Goded, I., Cescatti, A., Manca, G., and Seufert, G.: FLUXNET2015 IT-SRo San Rossore, FLUXNET [data set], https://doi.org/10.18140/flx/1440176 (last access: 1 November 2022), 2016a.

Gruening, C., Goded, I., Cescatti, A., and Pokorska, O.: FLUXNET2015 IT-Isp Ispra ABC-IS, FLUXNET [data set], https://doi.org/10.18140/flx/1440234 (last access: 1 November 2022), 2016b.

Hansen, B. U.: FLUXNET2015 GL-NuF Nuuk Fen, FLUXNET [data set], https://doi.org/10.18140/flx/1440222 (last access: 1 November 2022), 2016.

Haszpra, L.: Multi-decadal atmospheric carbon dioxide measurements in Hungary, central Europe, Atmos. Meas. Tech., 17, 4629–4647, https://doi.org/10.5194/amt-17-4629-2024, 2024.

Hayek, M. N., Wehr, R., Longo, M., Hutyra, L. R., Wiedemann, K., Munger, J. W., Bonal, D., Saleska, S. R., Fitzjarrald, D. R., and Wofsy, S. C.: A novel correction for biases in forest eddy covariance carbon balance, Agric. Forest Meteorol., 250-251, 90–101, https://doi.org/10.1016/j.agrformet.2017.12.186, 2018.

Hegarty, J., Draxler, R. R., Stein, A. F., Brioude, J., Mountain, M., Eluszkiewicz, J., Nehrkorn, T., Ngan, F., and Andrews, A.: Evaluation of Lagrangian Particle Dispersion Models with Measurements from Controlled Tracer Releases, J. Appl. Meteorol. Climatol., 52, 2623–2637, https://doi.org/10.1175/JAMC-D-13-0125.1, 2013.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. Roy. Meteor. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hinkle, R.: AmeriFlux FLUXNET-1F US-KS3 Kennedy Space Center (Salt Marsh), AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881586 (last access: 1 November 2022), 2022.

Hollinger, D.: AmeriFlux FLUXNET-1F US-Ho2 Howland Forest (West Tower), AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881581 (last access: 1 November 2022), 2022.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G.: Overview of the radiometric and biophysical performance of the MODIS vegetation indices, Remote Sens. Environ., 83, 195–213, https://doi.org/10.1016/S0034-4257(02)00096-2, 2002.

Huggins, D.: AmeriFlux FLUXNET-1F US-CF1 CAF-LTAR Cook East, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1832158 (last access: 1 November 2022), 2021.

Huggins, D.: AmeriFlux FLUXNET-1F US-CF3 CAF-LTAR Boyd North, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881574 (last access: 1 November 2022), 2022a.

Huggins, D.: AmeriFlux FLUXNET-1F US-CF4 CAF-LTAR Boyd South, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881575 (last access: 1 November 2022), 2022b.

Huggins, D.: AmeriFlux FLUXNET-1F US-CF2 CAF-LTAR Cook West, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881573 (last access: 1 November 2022), 2022c.

Jones, M. W., Andrew, R. M., Peters, G. P., Janssens-Maenhout, G., De-Gol, A. J., Ciais, P., Patra, P. K., Chevallier, F., and Le Quéré, C.: Gridded fossil CO2 emissions and related O2 combustion consistent with national inventories 1959–2018, Sci. Data, 8, 2, https://doi.org/10.1038/s41597-020-00779-6, 2021.

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., and Reichstein, M.: The FLUXCOM ensemble of global land-atmosphere energy fluxes, Sci. Data, 6, 74, https://doi.org/10.1038/s41597-019-0076-8, 2019.

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Goll, D. S., Haverd, V., Köhler,

P., Ichii, K., Jain, A. K., Liu, J., Lombardozzi, D., Nabel, J. E. M. S., Nelson, J. A., O'Sullivan, M., Pallandt, M., Papale, D., Peters, W., Pongratz, J., Rödenbeck, C., Sitch, S., Tramontana, G., Walker, A., Weber, U., and Reichstein, M.: Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach, Biogeosciences, 17, 1343–1365, https://doi.org/10.5194/bg-17-1343-2020, 2020.

Jung, M., Nelson, J., Migliavacca, M., El-Madany, T., Papale, D., Reichstein, M., Walther, S., and Wutzler, T.: Technical note: Flagging inconsistencies in flux tower data, Biogeosciences, 21, 1827–1846, https://doi.org/10.5194/bg-21-1827-2024, 2024.

Kaminski, T. and Heimann, M.: Inverse Modeling of Atmospheric Carbon Dioxide Fluxes, Science, 294, 259, https://doi.org/10.1126/science.294.5541.259a, 2001.

Kelley, H. J.: Gradient theory of optimal flight paths, Ars Journal, 30, 947–954, 1960.

Kendall, A., Gal, Y., and Cipolla, R.: Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7482–7491, IEEE, Salt Lake City, UT, USA, ISBN 978-1-5386-6420-9, https://doi.org/10.1109/CVPR.2018.00781, 2018.

Klatt, J., Schmid, H. P., Mauder, M., and Steinbrecher, R.: FLUXNET2015 DE-SfN Schechenfilz Nord, FLUXNET [data set], https://doi.org/10.18140/flx/1440219 (last access: 1 November 2022), 2016.

Knohl, A., Tiedemann, F., Kolle, O., Schulze, E.-D., Anthoni, P., Kutsch, W., Herbst, M., and Siebicke, L.: FLUXNET2015 DE-Lnf Leinefelde, FLUXNET [data set], https://doi.org/10.18140/flx/1440150 (last access: 1 November 2022), 2016.

Knox, S.: AmeriFlux FLUXNET-1F CA-DB2 Delta Burns Bog 2, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881564 (last access: 1 November 2022), 2022.

Kobayashi, H. and Suzuki, R.: FLUXNET2015 US-Prr Poker Flat Research Range Black Spruce Forest, FLUXNET [data set], https://doi.org/10.18140/flx/1440113 (last access: 1 November 2022), 2016.

Kondo, M., Ichii, K., Takagi, H., and Sasakawa, M.: Comparison of the data-driven top-down and bottom-up global terrestrial CO$_2$ exchanges: GOSAT CO$_2$ inversion and empirical eddy flux upscaling: Data-driven Terrestrial CO$_2$ Exchanges, J. Geophys. Res.-Biogeo., 120, https://doi.org/10.1002/2014JG002866, 2015.

Kondo, M., Patra, P. K., Sitch, S., Friedlingstein, P., Poulter, B., Chevallier, F., Ciais, P., Canadell, J. G., Bastos, A., Lauerwald, R., Calle, L., Ichii, K., Anthoni, P., Arneth, A., Haverd, V., Jain, A. K., Kato, E., Kautz, M., Law, R. M., Lienert, S., Lombardozzi, D., Maki, T., Nakamura, T., Peylin, P., Rödenbeck, C., Zhuravlev, R., Saeki, T., Tian, H., Zhu, D., and Ziehn, T.: State of the science in reconciling top-down and bottom-up approaches for terrestrial CO$_2$ budget, Global Change Biol., 26, 1068–1084, https://doi.org/10.1111/gcb.14917, 2020.

Kosugi, Y. and Takanashi, S.: FLUXNET2015 MY-PSO Pasoh Forest Reserve (PSO), FLUXNET [data set], https://doi.org/10.18140/flx/1440240 (last access: 1 November 2022), 2016.

Kotani, A.: FLUXNET2015 JP-SMF Seto Mixed Forest Site, FLUXNET [data set], https://doi.org/10.18140/flx/1440239 (last access: 1 November 2022), 2016a.

Kotani, A.: FLUXNET2015 JP-MBF Moshiri Birch Forest Site, FLUXNET [data set], https://doi.org/10.18140/flx/1440238 (last access: 1 November 2022), 2016b.

Kraft, B., Nelson, J. A., Walther, S., Gans, F., Weber, U., Duveiller, G., Reichstein, M., Zhang, W., Rußwurm, M., Tuia, D., Körner, M., Hamdi, Z., and Jung, M.: On the added value of sequential deep learning for the upscaling of evapotranspiration, Biogeosciences, 22, 3965–3987, https://doi.org/10.5194/bg-22-3965-2025, 2025.

Kurc, S.: AmeriFlux FLUXNET-1F US-SRC Santa Rita Creosote, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1871145 (last access: 1 November 2022), 2022.

Kusak, K., Sanchez, C., Szutu, D., and Baldocchi, D.: AmeriFlux FLUXNET-1F US-Snf Sherman Barn, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1854371 (last access: 1 November 2022), 2022.

Kutzbach, L.: AmeriFlux FLUXNET-1F AR-TF1 Rio Moat Bog, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1818370 (last access: 1 November 2022), 2021.

Law, B.: FLUXNET2015 US-Me3 Metolius-second Young Aged Pine, FLUXNET [data set], https://doi.org/10.18140/flx/1440080 (last access: 1 November 2022), 2016a.

Law, B.: FLUXNET2015 US-Me6 Metolius Young Pine Burn, FLUXNET [data set], https://doi.org/10.18140/flx/1440099 (last access: 1 November 2022), 2016b.

Law, B.: FLUXNET2015 US-Me1 Metolius – Eyerly Burn, FLUXNET [data set], https://doi.org/10.18140/flx/1440078 (last access: 1 November 2022), 2016c.

Law, B.: FLUXNET2015 US-Me5 Metolius-first Young Aged Pine, FLUXNET [data set], https://doi.org/10.18140/flx/1440082 (last access: 1 November 2022), 2016d.

Law, B.: FLUXNET2015 US-Me4 Metolius-old Aged Ponderosa Pine, FLUXNET [data set], https://doi.org/10.18140/flx/1440081 (last access: 1 November 2022), 2016e.

Law, B.: AmeriFlux FLUXNET-1F US-Me2 Metolius Mature Ponderosa Pine, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1854368 (last access: 1 November 2022), 2022.

Lee, H., Jung, M., Carvalhais, N., Trautmann, T., Kraft, B., Reichstein, M., Forkel, M., and Koirala, S.: Diagnosing modeling errors in global terrestrial water storage interannual variability, Hydrol. Earth Syst. Sci., 27, 1531–1563, https://doi.org/10.5194/hess-27-1531-2023, 2023.

Liddell, M. J.: FLUXNET2015 AU-Rob Robson Creek, Queensland, Australia, FLUXNET [data set], https://doi.org/10.18140/flx/1440203 (last access: 1 November 2022), 2016.

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.: A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian

Transport (STILT) model, J. Geophys. Res.-Atmos., 108, https://doi.org/10.1029/2002JD003161, 2003.

Lindauer, M., Steinbrecher, R., Wolpert, B., Mauder, M., and Schmid, H. P.: FLUXNET2015 DE-Lkb Lackenberg, FLUXNET [data set], https://doi.org/10.18140/flx/1440214 (last access: 1 November 2022), 2016.

Litvak, M.: AmeriFlux FLUXNET-1F US-Mpj Mountainair Pinyon-Juniper Woodland, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1832161 (last access: 1 November 2022), 2021.

Litvak, M.: AmeriFlux FLUXNET-1F US-Wjs Willard Juniper Savannah, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1871146 (last access: 1 November 2022), 2022.

Liu, H., Huang, M., and Chen, X.: AmeriFlux FLUXNET-1F US-Hn3 Hanford 100H Sagebrush, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881580 (last access: 1 November 2022), 2022.

Lohila, A., Aurela, M., Tuovinen, J.-P., Hatakka, J., and Laurila, T.: FLUXNET2015 FI-Jok Jokioinen, FLUXNET [data set], https://doi.org/10.18140/flx/1440159 (last access: 1 November 2022), 2016.

Luijkx, I. T., Chevallier, F., Roedenbeck, C., Niwa, Y., Liu, J., Feng, L., Palmer, P. I., Bowman, K., Peters, W., Tian, X., Piao, S., Zheng, B., Lloret, Z., Cozic, A., Jacobson, A. R., Yun, J., Byrne, B., Bloom, A., Jin, Z., Wang, Y., Zhang, H., Zhao, M., Wang, T., Ding, J., Liu, Z., Jiang, F., Ju, W., Yang, D., Chandra, N., and Patra, P.: Global CO2 gridded flux fields from 14 atmospheric inversions in GCB2023, ICOS [data set], https://doi.org/10.18160/4M52-VCRU, 2024.

Lund, M., Jackowicz-Korczyński, M., and Abermann, J.: FLUXNET2015 GL-ZaH Zackenberg Heath, FLUXNET [data set], https://doi.org/10.18140/flx/1440224 (last access: 1 November 2022), 2016a.

Lund, M., Jackowicz-Korczyński, M., and Abermann, J.: FLUXNET2015 GL-ZaF Zackenberg Fen, FLUXNET [data set], https://doi.org/10.18140/flx/1440223 (last access: 1 November 2022), 2016b.

Macfarlane, C., Lambert, P., Byrne, J., Johnstone, C., and Smart, N.: FLUXNET2015 AU-Gin Gingin, FLUXNET [data set], https://doi.org/10.18140/flx/1440199 (last access: 1 November 2022), 2016.

Manca, G. and Goded, I.: FLUXNET2015 IT-PT1 Parco Ticino Forest, FLUXNET [data set], https://doi.org/10.18140/flx/1440172 (last access: 1 November 2022), 2016.

Margolis, H. A.: FLUXNET2015 CA-Qfo Quebec – Eastern Boreal, Mature Black Spruce, FLUXNET [data set], https://doi.org/10.18140/flx/1440045 (last access: 1 November 2022), 2016.

Masarie, K. A., Peters, W., Jacobson, A. R., and Tans, P. P.: ObsPack: a framework for the preparation, delivery, and attribution of atmospheric greenhouse gas measurements, Earth Syst. Sci. Data, 6, 375–384, https://doi.org/10.5194/essd-6-375-2014, 2014.

Massman, B.: FLUXNET2015 US-GBT GLEES Brooklyn Tower, FLUXNET [data set], https://doi.org/10.18140/flx/1440118 (last access: 1 November 2022), 2016.

Massman, B.: AmeriFlux FLUXNET-1F US-GLE GLEES, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1871136 (last access: 1 November 2022), 2022.

Matamala, R.: FLUXNET2015 US-IB2 Fermi National Accelerator Laboratory- Batavia (Prairie Site), FLUXNET [data set], https://doi.org/10.18140/flx/1440072 (last access: 1 November 2022), 2016.

Matteucci, G.: FLUXNET2015 IT-Col Collelongo, FLUXNET [data set], https://doi.org/10.18140/flx/1440167 (last access: 1 November 2022), 2016.

McCaughey, H.: FLUXNET2015 CA-Gro Ontario – Groundhog River, Boreal Mixedwood Forest, FLUXNET [data set], https://doi.org/10.18140/flx/1440034 (last access: 1 November 2022), 2016.

Merbold, L., Rebmann, C., and Corradi, C.: FLUXNET2015 RU-Che Cherski, FLUXNET [data set], https://doi.org/10.18140/flx/1440181 (last access: 1 November 2022), 2016.

Metz, E.-M., Vardag, S. N., Basu, S., Jung, M., Ahrens, B., El-Madany, T., Sitch, S., Arora, V. K., Briggs, P. R., Friedlingstein, P., Goll, D. S., Jain, A. K., Kato, E., Lombardozzi, D., Nabel, J. E. M. S., Poulter, B., Séférian, R., Tian, H., Wiltshire, A., Yuan, W., Yue, X., Zaehle, S., Deutscher, N. M., Griffith, D. W. T., and Butz, A.: Soil respiration – driven CO$_2$ pulses dominate Australia's flux variability, Science, 379, 1332–1335, https://doi.org/10.1126/science.add7833, 2023.

Metz, E.-M., Vardag, S. N., Basu, S., Jung, M., and Butz, A.: Seasonal and interannual variability in CO$_2$ fluxes in southern Africa seen by GOSAT, Biogeosciences, 22, 555–584, https://doi.org/10.5194/bg-22-555-2025, 2025.

Meyer, W., Cale, P., Koerber, G., Ewenz, C., and Sun, Q.: FLUXNET2015 AU-Cpr Calperum, FLUXNET [data set], https://doi.org/10.18140/flx/1440195 (last access: 1 November 2022), 2016.

Meyers, T.: FLUXNET2015 US-LWW Little Washita Watershed, FLUXNET [data set], https://doi.org/10.18140/flx/1440077 (last access: 1 November 2022), 2016a.

Meyers, T.: FLUXNET2015 US-Goo Goodwin Creek, FLUXNET [data set], https://doi.org/10.18140/flx/1440070 (last access: 1 November 2022), 2016b.

Moncrieff, J., Malhi, Y., and Leuning, R.: The propagation of errors in long-term measurements of land-atmosphere fluxes of carbon and water, Global Change Biol., 2, 231–240, https://doi.org/10.1111/j.1365-2486.1996.tb00075.x, 1996.

Munassar, S., Rödenbeck, C., Koch, F.-T., Totsche, K. U., Gałkowski, M., Walther, S., and Gerbig, C.: Net ecosystem exchange (NEE) estimates 2006–2019 over Europe from a pre-operational ensemble-inversion system, Atmos. Chem. Phys., 22, 7875–7892, https://doi.org/10.5194/acp-22-7875-2022, 2022.

Munger, J. W.: FLUXNET2015 US-Ha1 Harvard Forest EMS Tower (HFR1), FLUXNET [data set], https://doi.org/10.18140/flx/1440071 (last access: 1 November 2022), 2016.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – A discussion of principles, J. Hydrol., 10, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Nelson, J. A., Walther, S., Gans, F., Kraft, B., Weber, U., Novick, K., Buchmann, N., Migliavacca, M., Wohlfahrt, G., Šigut, L., Ibrom, A., Papale, D., Göckede, M., Duveiller, G., Knohl, A.,

Hörtnagl, L., Scott, R. L., Dušek, J., Zhang, W., Hamdi, Z. M., Reichstein, M., Aranda-Barranco, S., Ardö, J., Op de Beeck, M., Billesbach, D., Bowling, D., Bracho, R., Brümmer, C., Camps-Valls, G., Chen, S., Cleverly, J. R., Desai, A., Dong, G., El-Madany, T. S., Euskirchen, E. S., Feigenwinter, I., Galvagno, M., Gerosa, G. A., Gielen, B., Goded, I., Goslee, S., Gough, C. M., Heinesch, B., Ichii, K., Jackowicz-Korczynski, M. A., Kloster-halfen, A., Knox, S., Kobayashi, H., Kohonen, K.-M., Korki-akoski, M., Mammarella, I., Gharun, M., Marzuoli, R., Mata-mala, R., Metzger, S., Montagnani, L., Nicolini, G., O'Halloran, T., Ourcival, J.-M., Peichl, M., Pendall, E., Ruiz Reverter, B., Roland, M., Sabbatini, S., Sachs, T., Schmidt, M., Schwalm, C. R., Shekhar, A., Silberstein, R., Silveira, M. L., Spano, D., Tages-son, T., Tramontana, G., Trotta, C., Turco, F., Vesala, T., Vincke, C., Vitale, D., Vivoni, E. R., Wang, Y., Woodgate, W., Yepez, E. A., Zhang, J., Zona, D., and Jung, M.: X-BASE: the first ter-restrial carbon and water flux products from an extended data-driven scaling framework, FLUXCOM-X, Biogeosciences, 21, 5079–5115, https://doi.org/10.5194/bg-21-5079-2024, 2024.

Network), N.: AmeriFlux FLUXNET-1F US-xBR NEON Bartlett Experimental Forest (BART), AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881598 (last access: 1 November 2022), 2022.

Nouvellon, Y.: FLUXNET2015 CG-Tch Tchizalamou, FLUXNET [data set], https://doi.org/10.18140/flx/1440142 (last access: 1 November 2022), 2016.

Novick, K. and Phillips, R.: AmeriFlux FLUXNET-1F US-MMS Morgan Monroe State Forest, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1854369 (last access: 1 November 2022), 2022.

Oikawa, P.: AmeriFlux FLUXNET-1F US-EDN Eden Land-ing Ecological Reserve, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1832159 (last access: 1 November 2022), 2021.

Olesen, J.: FLUXNET2015 DK-Fou Foulum, FLUXNET [data set], https://doi.org/10.18140/flx/1440154 (last access: 1 November 2022), 2016.

Ourcival, J.-M.: FLUXNET2015 FR-Pue Puechabon, FLUXNET [data set], https://doi.org/10.18140/flx/1440164 (last access: 1 November 2022), 2016.

O'Sullivan, M., Sitch, S., Friedlingstein, P., Luijkx, I. T., Peters, W., Rosan, T. M., Arneth, A., Arora, V. K., Chandra, N., Chevallier, F., Ciais, P., Falk, S., Feng, L., Gasser, T., Houghton, R. A., Jain, A. K., Kato, E., Kennedy, D., Knauer, J., McGrath, M. J., Niwa, Y., Palmer, P. I., Patra, P. K., Pongratz, J., Poulter, B., Rödenbeck, C., Schwingshackl, C., Sun, Q., Tian, H., Walker, A. P., Yang, D., Yuan, W., Yue, X., and Zaehle, S.: The key role of forest distur-bance in reconciling estimates of the northern carbon sink, Com-mun. Earth Environ., 5, 1–10, https://doi.org/10.1038/s43247-024-01827-4, 2024.

Pallandt, M. M. T. A., Kumar, J., Mauritz, M., Schuur, E. A. G., Virkkala, A.-M., Celis, G., Hoffman, F. M., and Göckede, M.: Representativeness assessment of the pan-Arctic eddy covari-ance site network and optimized future enhancements, Biogeo-sciences, 19, 559–583, https://doi.org/10.5194/bg-19-559-2022, 2022.

Palmer, P. I., Feng, L., Baker, D., Chevallier, F., Bösch, H., and Somkuti, P.: Net carbon emissions from African biosphere dom-

inate pan-tropical atmospheric CO2 signal, Nat. Commun., 10, 3344, https://doi.org/10.1038/s41467-019-11097-w, 2019.

Papale, D., Tirone, G., Valentini, R., Arriga, N., Belelli, L., Con-salvo, C., Dore, S., Manca, G., Mazzenga, F., Sabbatini, S., and Stefani, P.: FLUXNET2015 IT-Ro2 Roccarespampani 2, FLUXNET [data set], https://doi.org/10.18140/flx/1440175 (last access: 1 November 2022), 2016.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Reichstein, M., Ribeca, A., van Ingen, C., Vuichard, N., Zhang, L., Amiro, B., Ammann, C., Arain, M. A., Ardö, J., Arkebauer, T., Arndt, S. K., Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesder-fer, E., Marchesini, L. B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Boike, J., Bolstad, P. V., Bonal, D., Bonnefond, J.-M., Bowling, D. R., Bracho, R., Brodeur, J., Brümmer, C., Buch-mann, N., Burban, B., Burns, S. P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T. R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B. D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P. S., D'Andrea, E., da Rocha, H., Dai, X., Davis, K. J., Cinti, B. D., Grandcourt, A. d., Ligne, A. D., De Oliveira, R. C., Delpierre, N., Desai, A. R., Di Bella, C. M., Tommasi, P. d., Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrêne, E., Dunn, A., Dušek, J., Eamus, D., Eichel-mann, U., ElKhidir, H. A. M., Eugster, W., Ewenz, C. M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., Gharun, M., Gianelle, D., Gielen, B., Gioli, B., Gitelson, A., Goded, I., Goeckede, M., Goldstein, A. H., Gough, C. M., Goulden, M. L., Graf, A., Griebel, A., Gruening, C., Grünwald, T., Hammerle, A., Han, S., Han, X., Hansen, B. U., Hanson, C., Hatakka, J., He, Y., Hehn, M., Heinesch, B., Hinko-Najera, N., Hörtnagl, L., Hut-ley, L., Ibrom, A., Ikawa, H., Jackowicz-Korczynski, M., Janouš, D., Jans, W., Jassal, R., Jiang, S., Kato, T., Khomik, M., Klatt, J., Knohl, A., Knox, S., Kobayashi, H., Koerber, G., Kolle, O., Kosugi, Y., Kotani, A., Kowalski, A., Kruijt, B., Kurbatova, J., Kutsch, W. L., Kwon, H., Launiainen, S., Laurila, T., Law, B., Le-uning, R., Li, Y., Liddell, M., Limousin, J.-M., Lion, M., Liska, A. J., Lohila, A., López-Ballesteros, A., López-Blanco, E., Lou-bet, B., Loustau, D., Lucas-Moffat, A., Lüers, J., Ma, S., Mac-farlane, C., Magliulo, V., Maier, R., Mammarella, I., Manca, G., Marcolla, B., Margolis, H. A., Marras, S., Massman, W., Mas-tepanov, M., Matamala, R., Matthes, J. H., Mazzenga, F., Mc-Caughey, H., McHugh, I., McMillan, A. M. S., Merbold, L., Meyer, W., Meyers, T., Miller, S. D., Minerbi, S., Moderow, U., Monson, R. K., Montagnani, L., Moore, C. E., Moors, E., More-aux, V., Moureaux, C., Munger, J. W., Nakai, T., Neirynck, J., Nesic, Z., Nicolini, G., Noormets, A., Northwood, M., Nosetto, M., Nouvellon, Y., Novick, K., Oechel, W., Olesen, J. E., Our-cival, J.-M., Papuga, S. A., Parmentier, F.-J., Paul-Limoges, E., Pavelka, M., Peichl, M., Pendall, E., Phillips, R. P., Pilegaard, K., Pirk, N., Posse, G., Powell, T., Prasse, H., Prober, S. M., Rambal, S., Rannik, U., Raz-Yaseef, N., Rebmann, C., Reed, D., Dios, V. R. d., Restrepo-Coupe, N., Reverter, B. R., Roland, M., Sabbatini, S., Sachs, T., Saleska, S. R., Sánchez-Cañete, E. P., Sanchez-Mejia, Z. M., Schmid, H. P., Schmidt, M., Schneider, K., Schrader, F., Schroder, I., Scott, R. L., Sedlák, P., Serrano-Ortíz, P., Shao, C., Shi, P., Shironya, I., Siebicke, L., Šigut, L.,

Silberstein, R., Sirca, C., Spano, D., Steinbrecher, R., Stevens, R. M., Sturtevant, C., Suyker, A., Tagesson, T., Takanashi, S., Tang, Y., Tapper, N., Thom, J., Tomassucci, M., Tuovinen, J.-P., Urbanski, S., Valentini, R., van der Molen, M., van Gorsel, E., van Huissteden, K., Varlagin, A., Verfaillie, J., Vesala, T., Vincke, C., Vitale, D., Vygodskaya, N., Walker, J. P., Walter-Shea, E., Wang, H., Weber, R., Westermann, S., Wille, C., Wofsy, S., Wohlfahrt, G., Wolf, S., Woodgate, W., Li, Y., Zampedri, R., Zhang, J., Zhou, G., Zona, D., Agarwal, D., Biraud, S., Torn, M., and Papale, D.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Sci. Data, 7, 225, https://doi.org/10.1038/s41597-020-0534-3, 2020.

Pendall, E. and Griebel, A.: FLUXNET2015 AU-Cum Cumberland Plains, FLUXNET [data set], https://doi.org/10.18140/flx/1440196 (last access: 1 November 2022), 2016.

Peters, W., Krol, M. C., Van Der WERF, G. R., Houweling, S., Jones, C. D., Hughes, J., Schaefer, K., Masarie, K. A., Jacobson, A. R., Miller, J. B., Cho, C. H., Ramonet, M., Schmidt, M., Ciattaglia, L., Apadula, F., Heltai, D., Meinhardt, F., Di Sarra, A. G., Piacentino, S., Sferlazzo, D., Aalto, T., Hatakka, J., Ström, J., Haszpra, L., Meijer, H. A. J., Van Der Laan, S., Neubert, R. E. M., Jordan, A., Rodó, X., Morguí, J.-A., Vermeulen, A. T., Popa, E., Rozanski, K., Zimnoch, M., Manning, A. C., Leuenberger, M., Uglietti, C., Dolman, A. J., Ciais, P., Heimann, M., and Tans, P. P.: Seven years of recent European net terrestrial carbon dioxide exchange constrained by atmospheric observations, Global Change Biol., 16, 1317–1337, https://doi.org/10.1111/j.1365-2486.2009.02078.x, 2010.

Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C., van der Laan-Luijkx, I. T., and Zhang, X.: Global atmospheric carbon budget: results from an ensemble of atmospheric CO_2 inversions, Biogeosciences, 10, 6699–6720, https://doi.org/10.5194/bg-10-6699-2013, 2013.

Pilegaard, K. and Ibrom, A.: FLUXNET2015 DK-Eng Enghave, FLUXNET [data set], https://doi.org/10.18140/flx/1440153 (last access: 1 November 2022), 2016.

Posse, G., Lewczuk, N., Richter, K., and Cristiano, P.: FLUXNET2015 AR-Vir Virasoro, FLUXNET [data set], https://doi.org/10.18140/flx/1440192 (last access: 1 November 2022), 2016.

Poulter, B., Frank, D., Ciais, P., Myneni, R. B., Andela, N., Bi, J., Broquet, G., Canadell, J. G., Chevallier, F., Liu, Y. Y., Running, S. W., Sitch, S., and van der Werf, G. R.: Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle, Nature, 509, 600–603, https://doi.org/10.1038/nature13376, 2014.

Poveda, F. D., Ballesteros, A. L., Cañete, E. P. S., Ortiz, P. S., Jiménez, M. R. M., Priego, O. P., and Kowalski, A. S.: FLUXNET2015 ES-Amo Amoladeras, FLUXNET [data set], https://doi.org/10.18140/flx/1440156 (last access: 1 November 2022), 2016.

Randerson, J., Van Der Werf, G., Giglio, L., Collatz, G., and Kasibhatla, P.: Global Fire Emissions Database, Version 4.1 (GFEDv4), ORNL Distributed Active Archive Center [data set], https://doi.org/10.3334/ornldaac/1293, 2017.

Reverter, B. R., Perez-Cañete, E. S., and Kowalski, A. S.: FLUXNET2015 ES-Ln2 Lanjaron-Salvage Logging, FLUXNET [data set], https://doi.org/10.18140/flx/1440226 (last access: 1 November 2022), 2016a.

Reverter, B. R., Perez-Cañete, E. S., and Kowalski, A. S.: FLUXNET2015 ES-LgS Laguna Seca, FLUXNET [data set], https://doi.org/10.18140/flx/1440225 (last access: 1 November 2022), 2016b.

Rey-Sanchez, C., Wang, C., Szutu, D., Hemes, K., Verfaillie, J., and Baldocchi, D.: AmeriFlux FLUXNET-1F US-Bi2 Bouldin Island Corn, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1871135 (last access: 1 November 2022), 2022a.

Rey-Sanchez, C., Wang, C., Szutu, D., Shortt, R., Chamberlain, S., Verfaillie, J., and Baldocchi, D.: AmeriFlux FLUXNET-1F US-Bi1 Bouldin Island Alfalfa, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1871134 (last access: 1 November 2022), 2022b.

RI, I.: Ecosystem Final Quality (L2) Product in ETC-Archive Format – Release 2021-1, ICOS [data set], https://doi.org/10.18160/fzmy-pg92 (last access: 1 November 2022), 2021a.

RI, I.: Ecosystem Final Quality (L2) Product in ETC-Archive Format – Release 2022-1, ICOS [data set], https://doi.org/10.18160/pad9-hqhu (last access: 1 November 2022), 2022b.

Rödenbeck, C., Houweling, S., Gloor, M., and Heimann, M.: CO_2 flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport, Atmos. Chem. Phys., 3, 1919–1964, https://doi.org/10.5194/acp-3-1919-2003, 2003.

Rödenbeck, C., Zaehle, S., Keeling, R., and Heimann, M.: History of El Niño impacts on the global carbon cycle 1957–2017: a quantification from atmospheric CO_2 data, Phil. T. Roy. Soc. B, 373, 20170303, https://doi.org/10.1098/rstb.2017.0303, 2018.

Sabbatini, S., Arriga, N., Gioli, B., and Papale, D.: FLUXNET2015 IT-CA2 Castel d'Asso2, FLUXNET [data set], https://doi.org/10.18140/flx/1440231 (last access: 1 November 2022), 2016a.

Sabbatini, S., Arriga, N., Matteucci, G., and Papale, D.: FLUXNET2015 IT-CA3 Castel d'Asso 3, FLUXNET [data set], https://doi.org/10.18140/flx/1440232 (last access: 1 November 2022), 2016b.

Sabbatini, S., Arriga, N., and Papale, D.: FLUXNET2015 IT-CA1 Castel d'Asso1, FLUXNET [data set], https://doi.org/10.18140/flx/1440230 (last access: 1 November 2022), 2016c.

Sachs, T., Wille, C., Larmanou, E., and Franz, D.: FLUXNET2015 DE-Zrk Zarnekow, FLUXNET [data set], https://doi.org/10.18140/flx/1440221 (last access: 1 November 2022), 2016.

Saleska, S.: FLUXNET2015 BR-Sa1 Santarem-Km67-Primary Forest, FLUXNET [data set], https://doi.org/10.18140/flx/1440032 (last access: 1 November 2022), 2016.

Sanchez, C. R., Sturtevant, C., Szutu, D., Baldocchi, D., Eichelmann, E., and Knox, S.: FLUXNET2015 US-Tw4 Twitchell East End Wetland, FLUXNET [data set], https://doi.org/10.18140/flx/1440111 (last access: 1 November 2022), 2016.

Schaaf, C. and Wang, Z.: MCD43A2 MODIS/Terra+Aqua BRDF/Albedo Quality Daily L3 Global – 500 m V006, NASA

Land Processes Distributed Active Archive Center [data set], https://doi.org/10.5067/modis/mcd43a2.006 (last access: 1 November 2022), 2015a.

Schaaf, C. and Wang, Z.: MCD43A4 MODIS/Terra+Aqua BRDF/Albedo Nadir BRDF Adjusted Ref Daily L3 Global – 500 m V006, , NASA Land Processes Distributed Active Archive Center [data set], https://doi.org/10.5067/modis/mcd43a4.006 (last access: 1 November 2022), 2015b.

Schneider, K. and Schmidt, M.: FLUXNET2015 DE-Seh Selhausen, FLUXNET [data set], https://doi.org/10.18140/flx/1440217 (last access: 1 November 2022), 2016.

Schroder, I., Zegelin, S., Palu, T., and Feitz, A.: FLUXNET2015 AU-Emr Emerald, FLUXNET [data set], https://doi.org/10.18140/flx/1440198 (last access: 1 November 2022), 2016.

Scott, R.: FLUXNET2015 US-SRG Santa Rita Grassland, FLUXNET [data set], https://doi.org/10.18140/flx/1440114 (last access: 1 November 2022), 2016a.

Scott, R.: FLUXNET2015 US-SRM Santa Rita Mesquite, FLUXNET [data set], https://doi.org/10.18140/flx/1440090 (last access: 1 November 2022), 2016b.

Scott, R.: FLUXNET2015 US-Wkg Walnut Gulch Kendall Grasslands, FLUXNET [data set], https://doi.org/10.18140/flx/1440096 (last access: 1 November 2022), 2016c.

Scott, R.: FLUXNET2015 US-Whs Walnut Gulch Lucky Hills Shrub, FLUXNET [data set], https://doi.org/10.18140/flx/1440097 (last access: 1 November 2022), 2016d.

Shao, C.: FLUXNET2015 CN-Sw2 Siziwang Grazed (SZWG), FLUXNET [data set], https://doi.org/10.18140/flx/1440212 (last access: 1 November 2022), 2016a.

Shao, C.: FLUXNET2015 CN-Du3 Duolun Degraded Meadow, FLUXNET [data set], https://doi.org/10.18140/flx/1440210 (last access: 1 November 2022), 2016b.

Shi, P., Zhang, X., and He, Y.: FLUXNET2015 CN-Dan Dangxiong, FLUXNET [data set], https://doi.org/10.18140/flx/1440138 (last access: 1 November 2022), 2016.

Shortt, R., Hemes, K., Szutu, D., Verfaillie, J., and Baldocchi, D.: AmeriFlux FLUXNET-1F US-Sne Sherman Island Restored Wetland, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1871144 (last access: 1 November 2022), 2022.

Sigut, L., Havrankova, K., Jocher, G., Pavelka, M., and Janouš, D.: FLUXNET2015 CZ-BK2 Bily Kriz Grassland, FLUXNET [data set], https://doi.org/10.18140/flx/1440144 (last access: 1 November 2022), 2016.

Silveira, M.: AmeriFlux FLUXNET-1F US-ONA Florida Pine Flatwoods, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1832163 (last access: 1 November 2022), 2021.

Spano, D., Duce, P., Marras, S., Sirca, C., Arca, A., Zara, P., and Ventura, A.: FLUXNET2015 IT-Noe Arca Di Noe – Le Prigionette, FLUXNET [data set], https://doi.org/10.18140/flx/1440171 (last access: 1 November 2022), 2016.

Staebler, R.: AmeriFlux FLUXNET-1F CA-Cbo Ontario – Mixed Deciduous, Borden Forest Site, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1854365 (last access: 1 November 2022), 2022.

Sturtevant, C., Szutu, D., Baldocchi, D., Matthes, J. H., Oikawa, P., and Chamberlain, S. D.: FLUXNET2015 US-Myb Mayberry Wetland, FLUXNET [data set], https://doi.org/10.18140/flx/1440105 (last access: 1 November 2022), 2016.

Sturtevant, C., Verfaillie, J., and Baldocchi, D.: AmeriFlux FLUXNET-1F US-Tw2 Twitchell Corn, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881593 (last access: 1 November 2022), 2022.

Suyker, A.: FLUXNET2015 US-Ne3 Mead – Rainfed Maize-Soybean Rotation Site, FLUXNET [data set], https://doi.org/10.18140/flx/1440086 (last access: 1 November 2022), 2016a.

Suyker, A.: FLUXNET2015 US-Ne2 Mead – Irrigated Maize-Soybean Rotation Site, FLUXNET [data set], https://doi.org/10.18140/flx/1440085 (last access: 1 November 2022), 2016b.

Suyker, A.: AmeriFlux FLUXNET-1F US-Ne1 Mead – Irrigated Continuous Maize Site, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1871140 (last access: 1 November 2022), 2022.

Tagesson, T., Ardö, J., and Fensholt, R.: FLUXNET2015 SN-Dhr Dahra, FLUXNET [data set], https://doi.org/10.18140/flx/1440246 (last access: 1 November 2022), 2016.

Tang, Y., Kato, T., and Du, M.: FLUXNET2015 CN-HaM Haibei Alpine Tibet Site, FLUXNET [data set], https://doi.org/10.18140/flx/1440190 (last access: 1 November 2022), 2016.

Team, D. and Centre, I. E. T.: Drought-2018 Ecosystem Eddy Covariance Flux Product for 52 Stations in FLUXNET-Archive Format, ICOS [data set], https://doi.org/10.18160/yvr0-4898 (last access: 1 November 2022), 2020a.

Team, W. W.. and Centre, I. E. T.: Warm Winter 2020 Ecosystem Eddy Covariance Flux Product for 73 Stations in FLUXNET-Archive Format – Release 2022-1, ICOS [data set], https://doi.org/10.18160/2g60-zhak (last access: 1 November 2022), 2022b.

Torn, M.: FLUXNET2015 US-ARc ARM Southern Great Plains Control Site – Lamont, FLUXNET [data set], https://doi.org/10.18140/flx/1440065 (last access: 1 November 2022), 2016a.

Torn, M.: FLUXNET2015 US-ARb ARM Southern Great Plains Burn Site- Lamont, FLUXNET [data set], https://doi.org/10.18140/flx/1440064 (last access: 1 November 2022), 2016b.

Torn, M. and Dengel, S.: AmeriFlux FLUXNET-1F US-NGB NGEE Arctic Barrow, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1832162 (last access: 1 November 2022), 2021.

Tran, D. A., Gerbig, C., Rödenbeck, C., Panov, A., Prokushkin, A., Winderlich, J., Lavrič, J., Heimann, M., Seifert, T., Schultz, U., Schmidt, S., and Zaehle, S.: Zotino Tall Tower Observatory (ZOTTO) hourly CO_2 measurement dataset, Edmont [data set], https://doi.org/10.17617/3.YBPFG2, 2024a.

Tran, D. A., Gerbig, C., Rödenbeck, C., and Zaehle, S.: Interannual variations in Siberian carbon uptake and carbon release period, Atmos. Chem. Phys., 24, 8413–8440, https://doi.org/10.5194/acp-24-8413-2024, 2024b.

Upton, S., Reichstein, M., Gans, F., Peters, W., Kraft, B., and Bastos, A.: Constraining biospheric carbon dioxide fluxes by combined top-down and bottom-up approaches, Atmos. Chem. Phys., 24, 2555–2582, https://doi.org/10.5194/acp-24-2555-2024, 2024.

Upton, S., Reichstein, M., Peters, W., Botía, S., Nelson, J., Walther, S., Gans, F., Jung, M., and Haszpra, L.: Data from: Constraining a data-driven CO$_2$ flux model by ecosystem and atmospheric observations using atmospheric transport, Zenodo [data set], https://doi.org/10.5281/zenodo.17531189, 2026.

Valach, A., Shortt, R., Szutu, D., Eichelmann, E., Knox, S., Hemes, K., Verfaillie, J., and Baldocchi, D.: AmeriFlux FLUXNET-1F US-Tw1 Twitchell Wetland West Pond, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1832165 (last access: 1 November 2022), 2021.

Valach, A., Kasak, K., Szutu, D., Verfaillie, J., and Baldocchi, D.: AmeriFlux FLUXNET-1F US-Tw5 East Pond Wetland, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881595 (last access: 1 November 2022), 2022.

Valentini, R., Dore, S., Mazzenga, F., Sabbatini, S., Stefani, P., Tirone, G., and Papale, D.: FLUXNET2015 IT-Cpz Castelporziano, FLUXNET [data set], https://doi.org/10.18140/flx/1440168 (last access: 1 November 2022), 2016a.

Valentini, R., Nicolini, G., Stefani, P., De Grandcourt, A., and Stivanello, S.: FLUXNET2015 GH-Ank Ankasa, FLUXNET [data set], https://doi.org/10.18140/flx/1440229 (last access: 1 November 2022), 2016b.

Valentini, R., Tirone, G., Vitale, D., Papale, D., Arriga, N., Belelli, L., Dore, S., Manca, G., Mazzenga, F., Pegoraro, E., Sabbatini, S., and Stefani, P.: FLUXNET2015 IT-Ro1 Roccarespampani 1, FLUXNET [data set], https://doi.org/10.18140/flx/1440174 (last access: 1 November 2022), 2016c.

van der Laan-Luijkx, I. T., van der Velde, I. R., van der Veen, E., Tsuruta, A., Stanislawska, K., Babenhauserheide, A., Zhang, H. F., Liu, Y., He, W., Chen, H., Masarie, K. A., Krol, M. C., and Peters, W.: The CarbonTracker Data Assimilation Shell (CTDAS) v1.0: implementation and global carbon balance 2001–2015, Geosci. Model Dev., 10, 2785–2800, https://doi.org/10.5194/gmd-10-2785-2017, 2017.

Vivoni, E. and Perez-Ruiz, E.: AmeriFlux FLUXNET-1F US-Jo2 Jornada Experimental Range Mixed Shrubland, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881584 (last access: 1 November 2022), 2022.

Vourlitis, G., Dalmagro, H., De S. Nogueira, J., Johnson, M., and Arruda, P.: AmeriFlux FLUXNET-1F BR-Npw Northern Pantanal Wetland, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1881563 (last access: 1 November 2022), 2022.

Wagner-Riddle, C.: AmeriFlux FLUXNET-1F CA-ER1 Elora Research Station, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1832154 (last access: 1 November 2022), 2021.

Walther, S., Besnard, S., Nelson, J. A., El-Madany, T. S., Migliavacca, M., Weber, U., Carvalhais, N., Ermida, S. L., Brümmer, C., Schrader, F., Prokushkin, A. S., Panov, A. V., and Jung, M.: Technical note: A view from space on global flux towers by MODIS and Landsat: the FluxnetEO data set, Biogeosciences, 19, 2805–2840, https://doi.org/10.5194/bg-19-2805-2022, 2022.

Wang, H. and Fu, X.: FLUXNET2015 CN-Qia Qianyanzhou, FLUXNET [data set], https://doi.org/10.18140/flx/1440141 (last access: 1 November 2022), 2016.

Wohlfahrt, G., Hammerle, A., and Hörtnagl, L.: FLUXNET2015 AT-Neu Neustift, FLUXNET [data set], https://doi.org/10.18140/flx/1440121 (last access: 1 November 2022), 2016.

Wolf, S., Eugster, W., and Buchmann, N.: FLUXNET2015 PA-SPs Sardinilla-Pasture, FLUXNET [data set], https://doi.org/10.18140/flx/1440179 (last access: 1 November 2022), 2016a.

Wolf, S., Eugster, W., and Buchmann, N.: FLUXNET2015 PA-SPn Sardinilla Plantation, FLUXNET [data set], https://doi.org/10.18140/flx/1440180 (last access: 1 November 2022), 2016b.

Wood, J. and Gu, L.: AmeriFlux FLUXNET-1F US-MOz Missouri Ozark Site, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1854370 (last access: 1 November 2022), 2022.

Woodgate, W., Van Gorsel, E., and Leuning, R.: FLUXNET2015 AU-Tum Tumbarumba, FLUXNET [data set], https://doi.org/10.18140/flx/1440126 (last access: 1 November 2022), 2016.

Wu, D., Lin, J. C., Fasoli, B., Oda, T., Ye, X., Lauvaux, T., Yang, E. G., and Kort, E. A.: A Lagrangian approach towards extracting signals of urban CO$_2$ emissions from satellite observations of atmospheric column CO$_2$ (XCO2): X-Stochastic Time-Inverted Lagrangian Transport model ("X-STILT v1"), Geosci. Model Dev., 11, 4843–4871, https://doi.org/10.5194/gmd-11-4843-2018, 2018.

Yepez, E. and Garatuza, J.: AmeriFlux FLUXNET-1F MX-Tes Tesopaco, Secondary Tropical Dry Forest, AmeriFlux AMP [data set], https://doi.org/10.17190/amf/1832156 (last access: 1 November 2022), 2021.

Zhang, J. and Han, S.: FLUXNET2015 CN-Cha Changbaishan, FLUXNET [data set], https://doi.org/10.18140/flx/1440137 (last access: 1 November 2022), 2016.

Zhou, G. and Yan, J.: FLUXNET2015 CN-Din Dinghushan, FLUXNET [data set], https://doi.org/10.18140/flx/1440139 (last access: 1 November 2022), 2016.

Zona, D. and Oechel, W.: FLUXNET2015 US-Atq Atqasuk, FLUXNET [data set], https://doi.org/10.18140/flx/1440067 (last access: 1 November 2022), 2016a.

Zona, D. and Oechel, W.: FLUXNET2015 US-Ivo Ivotuk, FLUXNET [data set], https://doi.org/10.18140/flx/1440073 (last access: 1 November 2022), 2016b.