Constraining biospheric carbon dioxide fluxes by combined top-down and bottom-up approaches

Samuel Upton[1,2], Markus Reichstein[1], Fabian Gans[1], Wouter Peters[2,3], Basil Kraft[1], and Ana Bastos[1]

[1]Department of Biogeochemical Integration, Max Plank Institute of Biogeochemistry
[2]Environmental Sciences Group, Wageningen University, Wageningen, The Netherlands
[3]University of Groningen, Centre for Isotope Research, Groningen, The Netherlands

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

Abstract. While the growth rate of atmospheric CO_2 mole fractions can be measured with high accuracy, there are still large uncertainties its attribution to specific regions and diverse anthropogenic and natural sources and sinks. A major source of uncertainty is the net flux of carbon dioxide from the biosphere to the atmosphere, the Net Ecosystem Exchange (NEE). There are two major approaches to quantifying NEE: top-down approaches that typically use atmospheric inversions, and bottom-up
5   estimates which rely on process-based or data-driven models or inventories. Both top-down and bottom-up approaches have known strengths and limitations. Atmospheric inversions (e.g. those used in global carbon budgets) produce estimates of NEE that are consistent with the atmospheric CO_2 growth rate at regional and global scales, but are highly uncertain at smaller scales. Bottom-up data-driven models based on eddy-covariance measurements (e.g. FLUXCOM) match local observations of NEE and their spatial variability, but have difficulty in accurately upscaling to a reliable global estimate.

10   In this study, we propose to combine the two approaches to produce global NEE estimates, with the goal of capitalizing on each approach's strengths and mitigating their limitations. We do this by constraining the data-driven FLUXCOM model with regional estimates of NEE derived from an ensemble of atmospheric inversions from the Global Carbon Budget 2021. To do this, we need to overcome a series of scientific and technical challenges when combining information about diverse physical variables, which are influenced by different processes at different spatial and temporal scales. We design a modeling structure
15   that optimizes NEE by considering both the model's performance at the in-situ level, based on eddy-covariance measurements, and at the level of large regions, based on atmospheric inversion estimates of NEE and their uncertainty. This resulting "dual-constraint" data-driven flux model improves on information based on single constraints (top-down or bottom-up), producing robust locally-resolved and globally-consistent NEE spatio-temporal fields.

Compared to reference estimates of the global land sink from the literature, e.g. Global Carbon Budgets, our double-
20   constraint inferred global NEE shows a considerably smaller bias in global and tropical NEE compared to the underlying bottom-up data-driven model estimates (i.e., single constraint). The mean seasonality of our double-constraint inferred global NEE is also more consistent with the Global Carbon Budget and atmospheric inversions. At the same time, our model allows for more robustly spatially resolved NEE. The improved performance of the double-constraint model across spatial and temporal scales demonstrates the potential for adding top-down constraint to a bottom-up data-driven flux model.

1 Introduction

The annual growth of carbon dioxide in the atmosphere has been directly measured with high accuracy since 1957 (Keeling et al., 1989). However, attributing changes in atmospheric CO_2 to regional fluxes, and to the respective anthropogenic and natural sources and sinks is prone to uncertainties. One of the main sources of uncertainty is the land biosphere, which has large uncertainties both in the human-related and the natural flux components (Friedlingstein et al., 2022). Improved understanding of processes driving variations in the global carbon cycle requires, among other things, improved observational constraints on the net flux of carbon dioxide from the land surface to the atmosphere or Net Ecosystem Exchange (NEE), across local, regional, and global scales (Bastos et al., 2022; Ciais et al., 2022; Gaubert et al., 2019; Saeki and Patra, 2017; Thompson et al., 2016).

Two different approaches to constrain sources and sinks of carbon dioxide can be distinguished: top-down and bottom-up estimates (Crisp et al., 2022; Friedlingstein et al., 2022). Top-down approaches typically correspond to atmospheric inversions, which infer the surface fluxes over land and ocean based on observations of atmospheric CO_2 mole fractions based on a Bayesian inversion framework using prior estimates of the flux of carbon dioxide and an atmospheric transport model (Chevallier et al., 2005; Peylin et al., 2013; Crisp et al., 2022; Ciais et al., 2022). Bottom-up approaches typically rely on in-situ observations and/or remote-sensing data combined with statistical upscaling techniques, or process-based terrestrial biosphere and land-surface models to infer local to regional or global NEE (Jung et al., 2020; Kondo et al., 2020)

The strengths and limitations of each approach are largely inherited from the "point of view" of the system: Top-down approaches, using regional and global observations and meso- to global scale atmospheric transport, view the integral of fluxes from large areas. They produce reliable estimates of the magnitude and variability of latitudinal distribution of NEE, finding solutions that are in line with the global atmospheric growth rate (Gaubert et al., 2019; Friedlingstein et al., 2022). However, this aggregated view can compromise the local estimates of NEE, as the system adjusts sub-regional NEE to match the overall target, with estimates becoming increasingly uncertain for smaller regions (Ciais et al., 2010). Inverse models are rarely evaluated at the scale of individual ecosystem sites as the mismatch in spatial scales represented is simply too large.

Bottom-up approaches include a diversity of measurements, from small-scale direct observations at the leaf, plant, plot and ecosystem scale, to remote-sensing observations of relevant proxies (e.g. biomass, greenness) (Friedlingstein et al., 2022; Jung et al., 2020; Kondo et al., 2020). These approaches are sensitive to small-scale heterogeneity in the land surface and can provide information on the local distribution and magnitude of NEE. The FLUXCOM project (Jung et al., 2020) produced a comprehensive comparison of data-driven bottom-up approaches for upscaling terrestrial biosphere carbon dioxide and water fluxes based on eddy-covariance measurements. The FLUXCOM ensemble produced consistent spatial patterns of global NEE compared with process-based models (Jung et al., 2020), indicating that the model ensemble captured the relevant ecosystem-level processes. However, data-driven ecosystem-level flux models, including FLUXCOM, have a strong bias for NEE in the tropics compared to top-down estimates (Kondo et al., 2015; Jung et al., 2020), given that they depend on unevenly distributed eddy-covariance observations, which are particularly sparse in the tropics (Tramontana et al., 2016; Chu et al., 2017). Furthermore, micrometeorological conditions under the canopy in tropical forests can lead to data collection problems

at tropical towers due to low nighttime turbulence (Hayek et al., 2018; Fu et al., 2018; Jung et al., 2020) creating an incorrect learned relationship between driver variables and NEE. These two limitations result in global estimates of NEE that are far from other best estimates of NEE (Jung et al., 2020).

Previous studies have suggested that observations of atmospheric CO_2 could provide an additional constraint to bottom-up data-driven flux models (Jung et al., 2020; Anav et al., 2015). However, given the mismatch in spatial scales, processes and uncertainties between the two approaches (Ciais et al., 2022), even reconciling such estimates constitutes a challenge (Deng et al., 2021; Friedlingstein et al., 2022; Crisp et al., 2022; Ciais et al., 2022; Bastos et al., 2022). The two approaches also produce a different view of the flux, sensitive to different scales and processes of the land surface, such as fire and inland waters (Ciais et al., 2022). Thus it is not trivial to constrain a bottom-up data-driven flux model with a top-down view of atmospheric CO_2. Here, we aim to test the hypothesis, proposed in these previous studies, that a bottom-up data-driven flux model trained with a complementary constraint based on top-down atmospheric inversions can improve the estimates of regional and global land surface CO_2 fluxes. To do this, we first create a data-driven flux model analogue to a FLUXCOM member (Jung et al., 2020) trained on the observed NEE from eddy-covariance sites. Then, in parallel, we test the effect of adding an additional top-down constraint to the model's objective function used in NEE optimization. This top-down constraint is based on the regional integrals of NEE from an ensemble of atmospheric inversions. The addition of a top-down constraint to the bottom-up model requires solving a number of technical challenges to realistically link the two very different types of reference datasets in the objective function used to optimize NEE. We then evaluate the ability of the double-constrained model to estimate global and regional NEE, its spatial variability as well as temporal variability from seasonal to inter-annual time scales.

2 Data

2.1 Net Ecosystem Exchange

The carbon fluxes included in estimates from top-down and bottom up approaches are not exactly the same, as described in detail in Ciais et al. (2022). The primary observational constraint of top-down methods is atmospheric CO_2 observations, which are sensitive to nearly all carbon exchanges between the atmosphere and land ecosystems, as well as fluxes from inland water systems (lakes, rivers). In contrast, eddy-covariance flux measurements reflect carbon exchange across a smaller footprint of typically a few hundred meters, and their location is often selected to explicitly exclude the influence of fires and inland water systems. A meaningful comparison between fluxes derived from each thus requires adjusting for both fires, and inland water fluxes, as also discussed in Friedlingstein et al. (2022); Ciais et al. (2022); Deng et al. (2021). In this study, we account for these inherent differences as described in Section 2.3. From hereon, we refer to "NEE" as land carbon exchanges excluding inland water systems and fire fluxes.

2.2 Eddy-covariance site-level data

This study uses the same in-situ data as used in the FLUXCOM system in Jung et al. (2020), and Tramontana et al. (2016) (see the supplement from Tramontana et al. (2016) for a full list of included sites). The data-driven flux models are trained using meteorological observations and NEE data collected by globally distributed eddy-covariance towers in the La Thuile synthesis dataset of the and CarboAfrica network (Valentini et al., 2014) FLUXNET network (https://fluxnet.fluxdata.org/data/la-thuile-dataset/). Following Tramontana et al. (2016), driver variables are created using a set of remotely-sensed and meteorological data from the Moderate Resolution Imaging Spectroradiometer (MODIS) data (http://daac.ornl.gov/MODIS/), and the European Center for Medium-Range Weather Forecasting ERA5 atmospheric reanalysis dataset (https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5) (Tab. A2). At tower locations, the meteorological data is derived from the FLUXNET towers using ERA5 data for gap-filling following Tramontana et al. (2016), while the global dataset uses only ERA5 data. See Tramontana et al. (2016) for a full discussion of the handling of meteorological data. Furthermore, the FLUXCOM-RS+METEO V1 NEE ensemble data using ERA5 meteorological forcing from Jung et al. (2020) were used to generate the regional sparse linear models described below.

2.3 Atmospheric inversions

Atmospheric inversions use observed CO_2 mole fraction from in-situ measurement stations and flask network, or satellite retrievals of total column CO_2 (XCO2), with estimates of the non-biogenic fluxes to infer the exchange of carbon between the land, oceans and atmosphere. The land-atmosphere fluxes from atmosphere inversions correspond better to net biome productivity, or the total regional gain or loss of carbon from all processes, i.e. including the signal from fires and other disturbances, land use change and management, and river evasion (Ciais et al., 2022).

For the top-down constraint (referred to as "atmospheric"), we use the estimates of the land-atmosphere exchange from five models from the ensemble of atmospheric inversions from the Global Climate Budget (Friedlingstein et al., 2022, GCB2022) as described in Table 1. See Friedlingstein et al. (2022) for a full discussion of inversion systems. Only atmospheric inversions based on surface observations are used since they cover the study period used here, 2001-2017 (Table 1). All inversion results were provided on a common $1° \times 1°$ lat/lon grid and monthly temporal resolution.

The inversion estimates as provided by Friedlingstein et al. (2022) have been adjusted to rectify small differences in prescribed fossil fuel emissions, cement production and carbonation fluxes as well as lateral riverine CO_2 transport. We further subtracted fire emissions for each individual inversions at grid cell level. For the fire emission adjustment, we used the gridded fire fluxes from CAMS Global Fire Assimilation System (GFAS) (Di Giuseppe et al., 2018) without additional adjustment, which is consistent with the CarbonTracker Europe treatment of fire. The resulting fluxes thus represent land-ecosystem carbon exchange excluding inland waters and fires, comparable with NEE estimates based on upscale of eddy-covariance measurements.

While pixel-level NEE estimated by atmospheric inversions are known to be underconstrained (Ciais et al., 2010; Kaminski and Heimann, 2001), and are unlikely to provide robust constraints of NEE to train our model, atmospheric inversions produce

reliable estimates of the magnitude and variability of latitudinal distribution of NEE, finding solutions that are in line with the global atmospheric growth rate (Gaubert et al., 2019). Therefore, we aggregate the NEE from the inversions by a set of 18 very large regions consistent with the regions in the Regional Carbon Cycle Assessment and Processes-2 (RECCAP2) project (Tian et al., 2018). This aggregation leverages the growing consensus about the magnitude of global NEE from atmospheric

125 inversions (Gaubert et al., 2019) allowing for a "global" constraint from a set of smaller regional constraints which cover the land surface.

Note that the ensemble of inversions used here is not the source of the land sink estimate of the GCB2022, which is calculated as the residual land sink derived from other major independent terms in the global carbon budget (emissions from fossil fuels and industry (E_{FF}) + emissions from land-use change (E_{LUC}) - the ocean sink (S_{ocean}) - atmospheric growth rate (G_{ATM}),

130 and used here as reference for the global evaluation of our NEE estimates.

Table 1. Atmospheric inversions from the Global Climate Project 2022 (Friedlingstein et al. (2022)) and respective period covered and original references.

Inversion	Version	Date range	Reference
CAMS	v21r1	1979-2021	Chevallier et al. (2005)
CarbonTracker Europe (CTE)	v21r1	2001-2021	van der Laan-Luijkx et al. (2017)
Jena Carboscope (sEXTocNEET)	v2022	1957-2021	Rödenbeck et al. (2003), Rödenbeck et al. (2018)
UoE in situ	v6.1b	2001-2021	Feng et al. (2016), Palmer et al. (2019)
NISMON-CO2	v2022.1	1990-2021	Niwa et al. (2022)

3 Methods

This study is based on two models, illustrated in Fig. 1: a bottom-up data-driven model that upscales NEE from eddy-covariance measurements using a neural network with remote-sensing and meteorological predictors (hereafter referred to as the EC model) and a single objective function (training term 1 in Fig. 1), similar to the FLUXCOM model (Jung et al., 2019); and a

135 model that uses the same neural network structure but uses a dual objective function, which includes training term 1 (as in the EC model) and a second training term (training term 2) based on top-down regional constraint based on atmospheric inversions (hereafter referred to as the EC-ATM model). The optimization of NEE by EC-ATM is challenged by the spatial, temporal and physical units mismatch between the two training terms: training term 1 is based on daily in-situ NEE from eddy-covariance (in $gC/m^2/day$) and training term 2 is based on monthly NEE integrated over very large regions from atmospheric inversions

140 (in $PgC/month$). In order to link effectively across scales in the training of EC-ATM, we connect the neural network with a pre-computed statistical model that acts as a "bridge" between site-level NEE and regional integrals, allowing the neural network to learn efficiently from both bottom-up and top-down data streams.

In the next sections, we first present a description of the predictors used for NEE upscale (referred to as driver data (Section 3.1), then a description of the single- and double-constraint EC and EC-ATM models (Sections 3.2,3.3), including the statistical

145 approach proposed to bridge across scales in EC-ATM (Section 3.3.1), then we discuss the training approach for each model (Section 3.4) and finally the post-hoc analysis performed (Section 3.5).

3.1 Model driver data

We use the driver variables from the FLUXCOM-RS+METEO with ERA5 forcing ensemble as in Jung et al. (2020): Enhanced Vegetation Index (EVI), fraction of Absorbed Photosynthetically Active Radiation (fAPAR), Daytime Land Surface Temper-

150 ature (LST_{day}), Nighttime Land Surface Temperature (LST_{night}), the Medium Infrared Reflectance band (MIR), Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), extracted for each site and globally from MODIS, and the ERA5 variables incoming global radiation (R_g), top of atmosphere potential radiation (R_{pot}), Water Availability Index (WAI), and Air Temperature (T_{air}). The full set of drivers were constructed following Jung et al. (2020) and Tramontana et al. (2016), see Appendix table A2 for the driver formulation. These variables are computed globally and stored

155 by Plant Functional Type (PFT) derived from the MODIS Land Cover Type Yearly L3 Global 500m dataset Collection 5 (Friedl et al., 2010), and are reconstructed here by the percentage of the component PFTs in each pixel following the approach in Jung et al. (2019) and Tramontana et al. (2016). Two drivers, WAI and T_{air} are used at daily time step, while the other eight are constructed from the mean seasonal cycle (MSC) signal across the component variables. All are used at a 0.5°spatial resolution.

160 3.2 EC model description

The bottom-up data-driven flux model takes as input observations of meteorological and remotely-sensed drivers at a location, available from either eddy-covariance towers or satellite platforms, and outputs an inference of the NEE for that location. This bottom-up model consists of a feed-forward neural network, or a set of fully connected network layers, which we train using the standard gradient-based backpropagation algorithm (Kelley, 1960). The fully-connected layers consist of nodes or "neurons",

165 which are exposed to the output of all neurons in the previous layer. Non-linearity is introduced by passing each node output through a non-linear activation function. Our network is a set of three fully-connected layers with the ReLU activation function (Agarap, 2019). We provide a detailed illustration of the model architecture in Appendix Figure B1.

The bottom-up model (EC model) is trained using an objective function with one term which compares the model's inference of NEE and the observed NEE at the eddy-covariance site. The EC model is run and trained only on data from eddy-covariance

170 towers and co-located pixels. The EC model is identical to an ensemble member of the FLUXCOM system (Jung et al., 2020; Tramontana et al., 2016). For each experimental run we train the EC model independently, (Fig. 1, red lines), to provide a paired test of the impact of the additional atmospheric constraint.

3.3 EC-ATM model description

To improve regional and global upscaling performance, this study builds a second model (EC-ATM model), starting with the
same bottom-up model, to which we add a second term to the objective function comparing the NEE output, inferred at regional
scale using the regional models, with the integrals of regional NEE from an ensemble of atmospheric inversions (Fig. 1 orange
lines). To rule out effects of parameter initialization, we use the same initial state for the EC and EC-ATM neural networks.
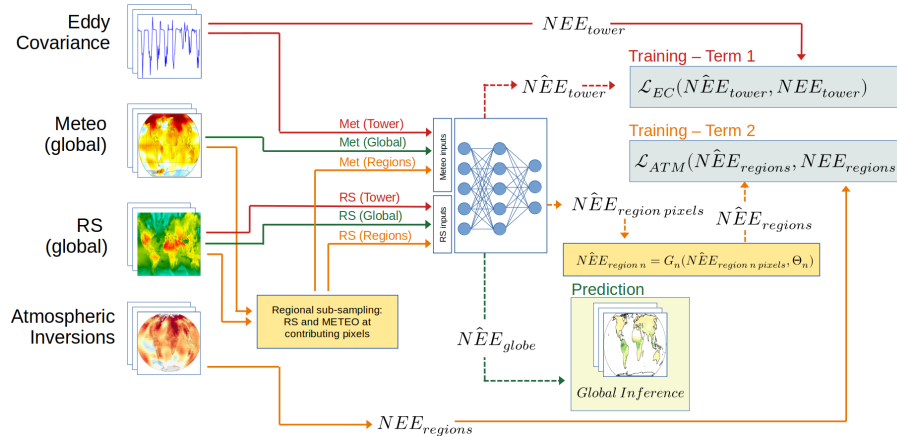
Figure 1. The EC model considers only the first term of the objective function (red lines). The EC-ATM model consists of the the bottom-up data-driven flux model (red lines) plus an additional constraint derived from atmospheric inversions (orange lines). In the first training pass, the neural network takes meteorological observations from eddy-covariance towers, along with Remotely Sensed (RS) data to create an inference of NEE, which is compared with the observed NEE in the first term of the objective function. In the second training pass, the neural network takes meteorological and RS variables at pre-selected pixels for each region. The inferred NEE at these pixels is fed into the regional bridge models to create inferences of regional NEE, which are compared with the regional integrals from an ensemble of atmospheric inversions. For global inference the neural network takes global meteorological data from ERA5 along with RS data to estimate NEE for all land pixels (green lines).

3.3.1 Statistical bridge-models

When calculating the atmospheric term of the objective function, running the bottom-up model for every land pixel and fully
calculating the global integral of NEE is too computationally intensive to train a data-driven model in a reasonable time frame.
This study solves this problem by using pre-computed statistical models to calculate fast approximations of the regional integral
based on a limited number of pixels for each region. We build this set of bridge models from the FLUXCOM RS+METEO
NEE results (Jung et al., 2020) by finding stable linear relationships between the NEE at fixed spatial locations within the
region with the regional integral of NEE in the same time period.

The bridge models are created using Least Absolute Shrinkage and Selection Operator (Lasso) regression. Lasso regression
extends Ordinary Least Squares (OLS) regression by adding a term to the objective function (Eq. 1) where N is the number of

observations, P is the number of independent variables, and y is the dependent variable. This term, is a weighted ℓ_1 norm, or the mean of the absolute values of the parameters of the linear model β, times the weighting term α. This additional term has the effect of driving some weights to zero as α is increased (Tibshirani, 1996).

$$Minimize: \sum_{i=1}^{N}(y_i - \sum_{j=1}^{P}x_{ij}\beta_j)^2 + \alpha\sum_{j=1}^{P}|\beta_j| \tag{1}$$

The dependent variable for the Lasso regression is the regional integral of the monthly ensemble mean NEE derived from the global NEE data from the FLUXCOM intercomparison, specifically the RS+METEO setup (Jung et al., 2020). The regional integrals are calculated as the sum of NEE for all non-zero pixels $p \in P_t$ in the region for time steps t in all times T (Eq. 2). The independent variables are the per-pixel monthly mean values of NEE, $p \in P_t$, for time step t. Using this formulation, the spatial logic of the regression is to find a stable set of weights relating the NEE at geographic locations within the region to the regional integral of NEE. This logic is then extended using Lasso regression (Eq. 3), where the reduction of weights in the model to zero creates a sparse solution. This can be interpreted as discovering a subset of geographic locations whose NEE are minimally sufficient to produce a high-quality estimate of the regional NEE from the FLUXCOM RS+METEO results. This quality is quantified as the R^2 value of the NEE inferred by the sparse Lasso model regressed against the regional integrals of NEE. The FLUXCOM RS+METEO data are thereafter only used for model evaluation.

$$NEE_t = \sum_{p=0}^{P} NEE_{tp} \tag{2}$$

$$Minimize: \sum_{t=0}^{T}(NEE_t - \sum_{p=0}^{P_s} NEE_{tp}\beta_p)^2 + \alpha\sum_{p=1}^{P_s}|\beta_p| \tag{3}$$

The full set of independent variables, or pixels in the region, are reduced to improve the stability of the regression, and to test the reliability of the technique by way of randomized trials. The reduced set of candidate pixels, $p \in P_s$, for region $s \in S_r$, is selected using stratified random sampling from a set of clusters generated using Dynamic Time Warping (DTW) (Tormene et al., 2008) as a metric of similarity, followed by spectral clustering. DTW is a method for comparing two time series which finds an minimum distance between them by allowing non-aligned time steps to be paired in the distance calculation as a similarity metric. This reduction attempts to reduce the number of candidate spatial locations while preserving the variance of the region. For each region a threshold of 0.95 R^2 is set, and the training loop iterates, reducing the α term, and increasing the number of non-zero weights in the region until the correlation threshold is reached using the training set. This result is tested using k-fold cross-validation. These resulting minimal set of spatial locations are the "contributing" pixels for the region. The locations of the contributing pixels and the corresponding weights and biases of the regional linear model constitute the statistical bridge model that will be used in model training.

The robustness of the sparse linear model was tested by 1000 runs with random stratified sampling within the discovered classes. The output shows stable spatial locations of contributing pixels within each region. A representative heat map of contributing pixels (Fig. B2) where the pixel value shows the log-scaled number of inclusions of that location in an iteration

of the Lasso regions, show that the Lasso approach repeatedly finds similar pixel locations, and will select spatial neighbors if the most advantageous pixels are not included in the randomization.

The stability of these spatial regimes indicate that there is a statistical link at the spatial resolution of the analysis between the contributing pixels and the regional integral of the EC model. The contributing pixels cover the range of the PFTs in the region (Fig. B3). The EC and EC-ATM models receive no PFT information during training, and the included PFT breakdown is based on the majority class at 0.5°resolution, not the available PFT information at the eddy-covariance sites, which is specific to the footprint of the EC tower.

3.4   Model Training

Each data-driven flux model is trained using a 10-fold cross-validation scheme, splitting the eddy-covariance observations by site into 10 equal subsets or "folds", holding out one fold per training cycle for validation, creating 10 model members. The composition of the folds is the same as those in Tramontana et al. (2016) for comparability. For the atmospheric inversion training data used by the EC-ATM model, two random years from the full 18 year set are held out for validation. There are insufficient years available for a fully independent set of test years outside the training and validation folds. We use the residual land sink from the CGB22 as independent data to test the model at global scale. Comparisons with the atmospheric inversions in results below are the same inversion data that is used in training. The reported results are for the ensemble mean across the 10 folds or members. For both the EC and EC-ATM models, the ensemble members are the data-driven flux model with the weights from the epoch with the best validation result for that fold. These members are used for a full forward run across the period 2001-2017. The results discussed below are calculated from these forward runs. Except for the very small number of pixels considered by the sparse linear models, these data are not considered during the training of the model. Unless otherwise stated, the results of the EC or EC-ATM model refer to the ensemble mean across the 10 members.

3.4.1   EC model training

At each training step, the EC model neural network f with parameters ω is run for a set of driver variables from eddy-covariance towers, and co-located RS measurements x_{batch} over a randomized subset of times and towers selected from all training observations (Eq. 4). The resulting inferences, $\widehat{NEE}$, are used to calculate the first term in the objective function, the loss at the EC tower locations, $\mathcal{L}_{\text{EC}}$, which is the Mean Squared Error (MSE) between the $\widehat{NEE}$, and the observed NEE at the tower locations in the training batch NEE_{obs} (Eq. 5).

$$\widehat{NEE} = f(\omega, x_{\text{batch}}) \tag{4}$$

$$\mathcal{L}_{EC} = MSE(NEE_{\text{obs}}, \widehat{NEE}) \tag{5}$$

3.4.2 EC-ATM model training

The EC-ATM model has two constraints. The first, $\mathcal{L}_{EC}$, is based on EC tower observations of NEE and driver variables measured at the EC tower sites and co-located RS data. This term is identical to the objective function of the EC model described above.

To create the second constraint $\mathcal{L}_{ATM}$, at each training step the EC-ATM model neural network f with parameters ω is run for a set of meteorological and RS driver variables collected at the "contributing" pixel locations in each region r in all regions R, for all days $d \in D$ in months m, notated $x_{d,r}$ (Eq. 6). These monthly values $\widehat{NEE_r}$ are then used to run the the statistical bridge-model for region r, with parameters (Θ_r, b_r), which produces an inference of the monthly integral of NEE for the month m, and region r (Eq. 7).

$$\widehat{NEE}_r = \frac{1}{D} \sum_{d=1}^{D} f(\omega, x_{d,r}) \tag{6}$$

$$\widehat{NEE}_{m,r} = \widehat{NEE}_r \cdot \Theta_r + b_r \tag{7}$$

This regional monthly estimate, $\widehat{NEE}_{r,m}$, is compared using the ℓ_3-norm with inversion-based estimates from inversions $a \in A$ of regional NEE, $a_{r,m}$, producing the monthly regional loss $\mathcal{L}_{r,m}$ (Eq. 8). The ℓ_3-norm was chosen because it improved the Interannual Variability (IAV) of the final NEE estimates.

$$\mathcal{L}_{r,m} = \sqrt[3]{\frac{1}{A} \sum_{a=1}^{A} |NEE_{a,r,m} - \widehat{NEE}_{r,m}|^3} \tag{8}$$

This $\mathcal{L}_{r,m}$ term is normalized by the range (Eq. 9) of the regional NEE from atmospheric inversions A for that month $range_{r,m}$ (Eq. 10). This reduces the weight of the loss where the atmospheric inversion ensemble has a higher uncertainty. The weighted losses are averaged for all months in M creating a regional loss term $\mathcal{L}_{ATM,r}$. The area-weighted average of the regional losses creates an atmospheric loss term $\mathcal{L}_{ATM}$ (Eq. 11). The normalization by the range of the inversions accounts for the uncertainty between the inversion systems. It acts to reduce the weight of the error term in proportion to this relative uncertainty for each step.

$$range_{r,m} = max(NEE_{A,r,m}) - min(NEE_{A,r,m}) \tag{9}$$

$$\mathcal{L}_{ATM,r} = \frac{1}{M} \sum_{m=1}^{M} \left(\frac{\mathcal{L}_{r,m}}{1 + range_{r,m}} \right) \tag{10}$$

$$\mathcal{L}_{ATM} = \sum_{r=1}^{R} \mathcal{L}_{ATM,r} \times \frac{land\,area_r}{land\,area_{globe}} \tag{11}$$

The two terms of the objective function, $\mathcal{L}_{EC}$ and $\mathcal{L}_{ATM}$ are combined using an empirically learned weighting scheme (Kendall et al., 2018), which learns the appropriate relative weights for the set of losses. To do this, the method adds a parameter to the learned weights of the data-driven model that estimates the task-dependent, homoscedastic uncertainty for each of the different terms of the objective function, which is dependent on the inherent noise in the data, rather than the scale or

quality of the inputs. This term is an estimate of the variance of the error term for each component loss of the objective function over all training steps. For the EC-ATM model these parameters, σ_{EC}^2 and σ_{ATM}^2, are added to the model training weights, and updated by the regular backpropagation step of the neural network training. The σ_{LOSS}^2 parameter is used to create a weighting term w_{LOSS} (Eq. 12) and a regularization term s_{LOSS} (Eq. 12) for each component term of the objective function, $LOSS \in [EC, ATM]$. These are then combined to provide a learned estimate of the total loss, balanced by the learned uncertainty of the terms (Eq. 14).

The total loss of the EC-ATM model is then:

$$w_{EC} = \frac{1}{2\sigma_{EC}^2}, \quad w_{ATM} = \frac{1}{2\sigma_{ATM}^2} \tag{12}$$

$$s_{EC} = \log\sqrt{\sigma_{EC}^2}, \quad s_{ATM} = \log\sqrt{\sigma_{ATM}^2} \tag{13}$$

$$\mathcal{L}_{total} = (w_{EC} \times \mathcal{L}_{EC}) + (w_{ATM} \times \mathcal{L}_{ATM}) + s_{EC} + s_{ATM} \tag{14}$$

3.5 Post-Hoc Analysis

After training any specific model, we carefully checked the validity of our assumptions, and the appropriateness of using bridge models. A known limitation of this method is the instability caused by large changes in the learned spatial pattern of NEE during training. These changes can lead to a decoupling between the model response and the NEE data from FLUXCOM RS+METEO V1 underlying the bridge models. This means the $\mathcal{L}_{ATM}$ is reduced, while the difference between the full integral of inferred NEE moves away from the inversion estimate. This decoupling destabilizes the learning process because the regional integral of NEE that is encoded in the regional bridge model is no longer valid for the output of the model being trained. This leads to a situation where certain random states of model initialization create unrealistic model results. We conduct a post-hoc test of the relationship between the regional sparse linear bridge models and the calculated integrals by applying the regional sparse linear models over the contributing pixel locations within the output of our training runs. This allows us to diagnose these changes in the learned spatial pattern of NEE during training.

3.6 One-against-many Inversion Sensitivity Analysis

We assume that the spread across the inversion estimates of regional NEE at each training set allows for improved NEE constraints by providing a measure of their uncertainty. In order to evaluate how the EC-ATM NEE estimates depend on this uncertainty constraint, we performed a sensitivity analysis where we trained several models with the atmospheric constraint coming from either one inversion (zero spread), two randomly selected inversions, or three randomly selected inversions, in addition to the standard setup with five inversions. The goal of this analysis is to evaluate how NEE from the EC-ATM model trained with these limited subsets differs from NEE calculated using the full ensemble of inversions. This allows to better understand how our use of an ensemble of inversions in training, and the uncertainty normalization strategy influence the use of information in the model.

4 Results

4.1 Mean annual fluxes and inter-annual variability

305 We first compare the global NEE estimates from the EC and the EC-ATM models (Fig. 2) with the residual estimate of the land sink from the Global Carbon Budget (GCB22) (Friedlingstein et al., 2022). The addition of an atmospheric constraint to a data-driven flux model (i.e. the EC-ATM model) leads to a global estimate of NEE (-3.14±1.75 PgC year^{-1} (± 1 σ)) that is closer to the current best estimates from GCB22 (-2.99±0.9 PgC year^{-1} (± 1 σ)) than the model only based on eddy-covariance flux data (EC model, -20.28±1.75 PgC year^{-1}). The GCB22 estimate represents an independent view of the sink magnitude as

310 it is calculated from as the residual of other major independent terms in the global carbon budget ($E_{FF} + E_{LUC} - S_{ocean} - G_{ATM}$) which are not used in our approach. Moreover, it agrees well with estimates of the land sink from other recent studies (Table C1). The EC-ATM annual mean NEE has an RMSE of 0.91 PgC year^{-1}(compared to GCB22) for the period 2001-2017, compared to an RMSE of 17.32 PgC year^{-1} for the EC model. The EC model estimates are consistent with the original FLUXCOM RS+METEO V1 estimates (Jung et al., 2020) (Fig. 2 yellow line).

315 The GCB22 results are not available at the RECCAP2 regional level, so to assess the performance of the models at regional scale we use the ensemble of atmospheric inversions for comparison. We acknowledge that the estimates are not fully independent, since they are based on the same data used to train our model, but we expect our training approach (see section 3.4 above) to make NEE estimates from the EC-ATM model sufficiently distinct from the ensemble mean of atmospheric inversions which is not used in directly the training.

320 Figure 2b shows that the mean anomaly of NEE estimated by EC-ATM model largely captures the sign of the atmospheric inversion and GCB22 anomalies, but the IAV is still underestimated by both the EC-ATM and EC models, which is a persistent issue with the FLUXCOM approach (see Jung et al. (2020) for full discussion) as well as a general issue with statistical learning. The EC-ATM annual mean is strongly correlated with the detrended standardized annual anomalies of the atmospheric inversion ensemble mean (R^2 of 0.67), producing very similar results to the FLUXCOM RS+METEO results (Fig. 2B).

325 In annual regional results (Table 2), the EC-ATM model generally outperforms the EC and FLUXCOM RS+METEO models in the tropics, with more mixed results in the extratropical regions. There is an improvement in RMSE for most regions. The RMSE for Brazil improves over the FLUXCOM RS+METEO, going from 3.89 PgC year^{-1} to 0.13 PgC year^{-1}, and Europe improves from 0.68 to 0.21 PgC year^{-1}. The EC-ATM model appears to have lower performance for boreal regions compared with FLUXCOM. Both Canada (CAN) and Russia (RUS) having higher RMSE compared with the inversion ensemble mean

330 (0.63 PgC year^{-1} for the EC-ATM compared with 0.18 PgC year^{-1} for FLUXCOM in RUS). The correlation does not show clear or systematic improvement. The Pearson's R for most regions are largely similar (see result for the USA, Equatorial Africa (EQAF) and Southeast Asia (SEAS)). This is consistent with the an overall low performance in capturing the IAV signal with data-driven flux models. Of note is the fact that while the boreal regions have higher errors, they have stronger correlations (0.53 PgC year^{-1} for the EC-ATM compared with -0.41 PgC year^{-1} for the EC in RUS), indicating a larger magnitude of error,

335 but better estimation of the monthly ecosystem dynamics.
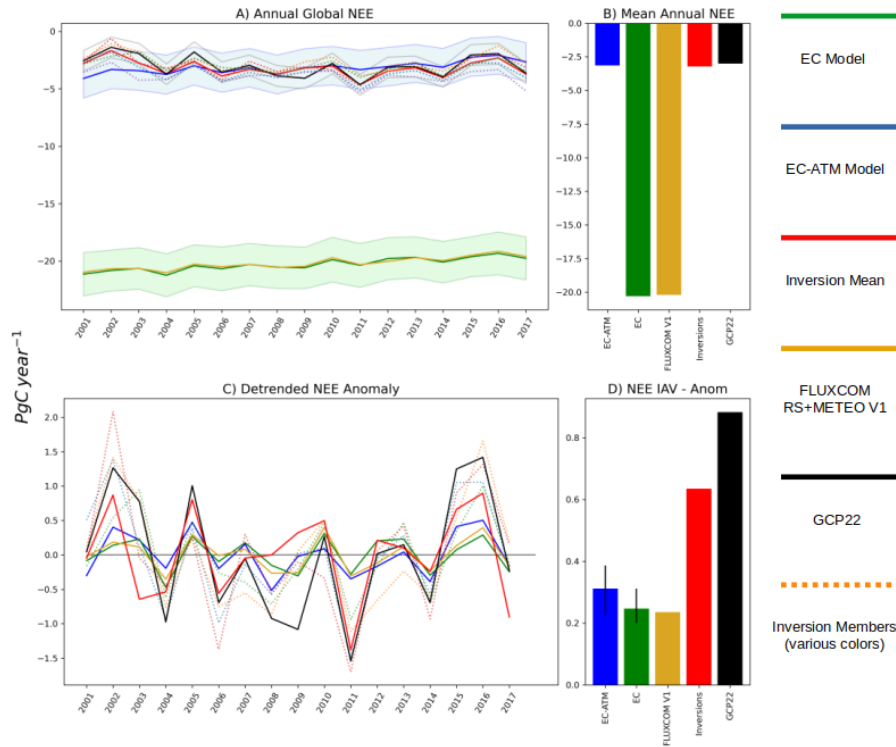
Figure 2. Panel (A) is the global annual NEE in PgC year⁻¹. The blue line is the EC-ATM ensemble mean across the 10 members The blue shaded area is the EC-ATM uncertainty across the ensemble ($\pm 1\sigma$). The green line is the EC ensemble mean across the 10 members. The green shaded area is the EC-ATM uncertainty across the ensemble ($\pm 1\sigma$). The red line is the ensemble mean of the atmospheric inversions. The individual inversions are shown in dotted lines. The black line is the GCP22 residual land sink. The grey shaded area is the published GCB22 uncertainly. The yellow line is the ensemble mean of FLUXCOM RS+METEO V1. (B) is the annual mean across 2001-2017. (C) shows the detrended anomalies in PgC year⁻¹. (D) shows the magnitude of IAV, estimated as the standard deviation of detrended annual anomalies. The bars on the EC and EC-ATM bars indicate the spread of IAV across the 10 members.

4.2 Monthly and seasonal variability

We evaluate the performance of the EC and EC-ATM models in capturing the overall temporal variability in NEE by estimating the normalized Nash Sutcliffe model efficiency (nNSE) metric for regional monthly NEE values over all years (2001-2017) (Fig. 3). Normalized Nash Sutcliffe model efficiency is a transformation of standard Nash Sutcliffe metric which assesses 340 the predictive skill of a model in regard to a reference set of values, in this instance, the monthly regional integrals of the atmospheric inversion ensemble. The normalization transforms the range of the metric from $(-\infty, 1)$ to $(0, 1)$. An nNSE of 1.0 represents perfect skill where the EC-ATM or EC perfectly reproduce this reference. An nNSE of 0 represents no skill. An nNSE of 0.5 is where the model predicts the reference better than repeating the annual regional mean integrals of the

Table 2. Results of annual NEE aggregated by regions. Pearson's correlation coefficient (R) of the annual integral of NEE from the inversion ensemble mean with the machine-learned model estimates from EC-ATM, EC, and FLUXCOM. The final row is the Pearson's R and RMSE globally for the three models compared with the GCB22 results. Units are PgC year^{-1}. Bold numbers denote the best performing model by region and metric. Country/region abbreviations are expanded in Appendix table A1

	Pearson's R – Annual			RMSE		
	EC-ATM	**EC**	**FLUXCOM V1**	**EC-ATM**	**EC**	**FLUXCOM V1**
GLOBAL	**0.280**	0.054	0.080	**0.7502**	16.8872	17.3236
USA	0.502	0.519	**0.543**	**0.2303**	0.8030	1.0836
CAN	**0.325**	-0.463	-0.236	0.3970	0.3747	**0.0627**
CAM	**-0.067**	-0.085	-0.149	**0.2236**	0.6598	0.7898
NSA	-0.360	**-0.140**	-0.253	**0.0969**	1.7467	1.4909
BRA	**0.358**	0.221	0.186	**0.1319**	3.8447	3.8912
SSA	-0.037	-0.094	**0.007**	**0.2021**	1.9080	1.8467
EU	-0.068	0.187	**0.191**	**0.2074**	0.6336	0.6805
NAF	0.557	0.616	**0.694**	**0.1900**	0.4362	0.2438
EQAF	**0.749**	0.725	0.740	**0.3876**	2.8844	2.8576
SAF	**0.780**	0.734	0.621	**0.1352**	0.5242	1.0036
RUS	**0.534**	-0.353	-0.409	0.6273	1.1558	**0.1809**
CAS	-0.379	-0.258	**-0.167**	0.1326	0.7133	**0.0759**
MIDE	0.254	0.405	**0.442**	0.2277	0.5117	**0.1585**
CHN	0.084	**0.297**	0.292	**0.1103**	1.0917	0.8059
KAJ	0.185	0.173	**0.216**	**0.0147**	0.0798	0.1167
SAS	**0.262**	0.066	0.039	**0.1384**	0.9604	0.3685
SEAS	-0.227	-0.263	**-0.221**	**0.4350**	2.0118	2.0409
OCE	0.499	0.599	**0.508**	0.4094	0.3703	**0.1037**
GCB22	**0.303**	0.115	0.138	**0.914**	17.316	17.228

atmospheric inversion ensemble. Figure 3 shows that in almost all regions the EC-ATM model is better able to predict the
345 atmospheric inversions than the EC model.

The mean seasonal cycle (MSC) of the global NEE estimated by the EC-ATM model shows a clear adjustment towards the atmospheric inversion ensemble mean and away from the estimates from the EC model and FLUXCOM RS+METEO results, consistent with the rationale to use a double-constraint approach (Fig. 4). In extratropical regions (e.g. EU in Fig. 4) the MSC of the EC-ATM is very close to the FLUXCOM RS+METEO and inversion ensemble mean results. But globally, and in tropical
350 regions the MSC shows a meaningful improvement (Table 3). The range of the global MSC (Fig. 4) is closer to the atmospheric inversion ensemble mean, with 3.66 PgC and 3.88 PgC respectively, compared to 2.24 PgC in the EC model. The EC-ATM MSC has an RMSE of 0.13 PgC compared with 1.54 PgC for FLUXCOM RS+METEO and 1.5 PgC for the EC model. The seasonality of the EC-ATM inferred flux is also closer to that of the atmospheric inversion ensemble mean, although the EC-

Table 3. Results of monthly NEE aggregated by regions: Pearson's R and RMSE of the monthly time-series of regional/global integrals over the period 2001-2017 and the corresponding monthly NEE from atmospheric inversions, and the Pearson's R of the regional/global Mean Seasonal Cycle (MSC) of NEE, the RMSE of the MSC relative to the inversion mean and the model MSC. FLUXCOM refers to the RS+METEO V1 product (Jung et al., 2020). Bold numbers denote the best performing model by region and metric. Country/region abbreviations are expanded in Appendix table A1

.

	Pearson's R – monthly integrals			RMSE – monthly integrals			Pearson's R – MSC			RMSE – MSC		
	EC-ATM	EC	FLUXCOM	EC-ATM	EC	FLUXCOM	EC-ATM	EC	FLUXCOM	EC-ATM	EC	FLUXCOM
GLOBAL	**0.991**	0.969	0.976	**0.176**	1.502	1.542	**0.996**	0.976	0.982	**0.125**	1.496	1.537
USA	**0.989**	0.971	0.974	**0.036**	0.090	0.105	**0.998**	0.979	0.982	**0.022**	0.086	0.101
CAN	**0.991**	0.949	0.974	**0.048**	0.083	0.049	**0.997**	0.958	0.981	**0.044**	0.080	0.045
CAM	**0.658**	0.557	0.607	**0.022**	0.061	0.071	**0.861**	0.685	0.751	**0.020**	0.060	0.070
NSA	**0.133**	-0.007	-0.023	**0.018**	0.148	0.127	**0.197**	-0.072	-0.080	**0.012**	0.147	0.126
BRA	**0.787**	-0.028	0.040	**0.055**	0.354	0.350	**0.893**	-0.051	0.024	**0.043**	0.352	0.348
SSA	**0.751**	0.388	0.387	**0.032**	0.176	0.173	**0.869**	0.431	0.431	**0.025**	0.175	0.172
EU	0.988	0.986	**0.989**	0.034	0.059	0.061	**0.997**	0.994	0.997	**0.026**	0.055	0.058
NAF	**0.929**	0.809	0.902	**0.031**	0.058	0.033	**0.959**	0.834	0.926	**0.027**	0.056	0.030
EQAF	**0.675**	0.186	0.253	**0.057**	0.251	0.247	**0.783**	0.173	0.248	**0.046**	0.249	0.245
SAF	**0.920**	0.600	0.665	**0.037**	0.102	0.117	**0.982**	0.628	0.698	**0.025**	0.099	0.114
RUS	**0.993**	0.931	0.970	**0.074**	0.225	0.138	**0.999**	0.939	0.977	**0.059**	0.220	0.130
CAS	**0.936**	0.835	0.655	**0.039**	0.069	0.054	**0.957**	0.856	0.698	**0.036**	0.067	0.052
MIDE	0.913	0.882	**0.919**	0.025	0.046	**0.020**	0.926	0.893	**0.931**	0.025	0.046	**0.019**
CHN	0.969	0.973	**0.973**	**0.035**	0.104	0.073	0.987	**0.990**	0.990	**0.025**	0.101	0.070
KAJ	**0.962**	0.937	0.939	**0.004**	0.009	0.012	**0.984**	0.960	0.959	**0.003**	0.008	0.011
SAS	**0.803**	0.184	0.607	**0.028**	0.111	0.046	**0.896**	0.208	0.680	**0.020**	0.109	0.042
SEAS	**0.273**	0.204	0.204	**0.050**	0.173	0.174	**0.436**	0.312	0.300	**0.042**	0.170	0.172
OCE	0.199	0.223	**0.240**	0.044	0.047	**0.036**	0.289	0.263	**0.310**	0.037	0.041	**0.030**

ATM model appears to underestimate the source during the Northern Hemisphere winter and conversely overestimate it in the Northern Hemisphere early spring (Appendix Fig. C1).

The EC-ATM model shows an improvement in the global RMSE of monthly NEE from 1.54 PgC/month for the EC mean to 0.13 PgC/month for the EC-ATM model mean (Fig. 3). The regional monthly performance of the EC-ATM model shows improvements in both the monthly time series and MSC for almost all regions (Tab. 3). In extratropical regions the improvements are small. The EC-ATM NEE for Canada (CAN) has a monthly Pearson's R of 0.991 compared with 0.969 for the EC model, and .949 for FLUXCOM RS+METEO. However in tropical regions the change is much larger. The EC-ATM NEE for Brazil (BRA) has a monthly Pearson's R of 0.787 compared with 0.028 for the EC model and 0.040 for FLUXCOM. Regions where the EC or FLUXCOM outperform the EC-ATM results are all in regions where all three models perform very similarly.

In appendix Figure C1 impact of the atmospheric information on the seasonality of the MSC is quite clear. In tropical regions (e.g. Central America (CAM), South Asia (SAS), BRA) the EC-ATM result has a very different seasonality than the EC and

365   FLUXCOM RS+METEO results, including different ranges and means. The correspondence between the inversion ensemble mean and EC-ATM MSC is close, but the differences reflect the balance between eddy-covariance and atmospheric information during model training, and the higher uncertainty between inversions in tropical regions. The difference between the EC model MSC and the FLUXCOM RS+METEO MSC can be attributed to different initial model states. The EC model is an analog for a member of the FLUXCOM RS+METEO ensemble, the mean of which is used here for comparison. In the extratropical

370   regions, the estimates of NEE MSC by EC and EC-ATM are very similar in range and seasonality.

These results indicate that additional atmospheric constraints are indeed reflected in the EC-ATM model at seasonal and continental scales, i.e., the scales where atmospheric data are most informative.
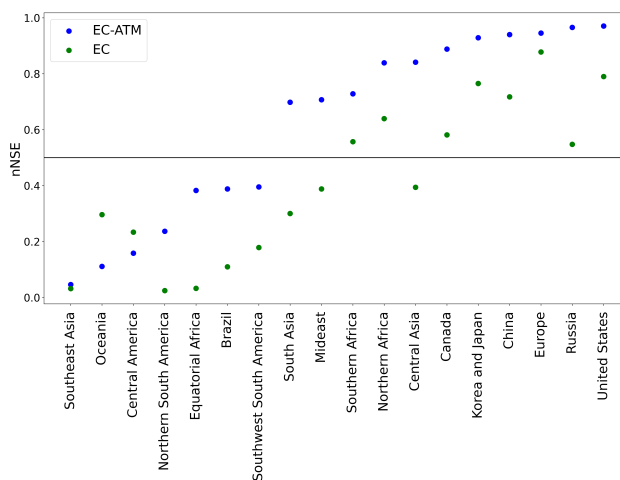
Figure 3. Normalized Nash-Sutcliffe model efficiency over all regions ordered by EC-ATM performance. An nNSE of 1.0 represents perfect skill where the EC-ATM or EC perfectly reproduce the monthly regional integrals of the atmospheric inversion ensemble. An nNSE of 0 represents no skill. An nNSE of 0.5 is where the model predicts the reference better than repeating the annual mean of the atmospheric inversion ensemble.

Figure 4 shows two selected regions: (1) Brazil which is representative of tropical regions with sparse EC observations and potential systematic bias in the EC data, and (2) Europe which is representative of extratropical regions with a dense

375   EC network. In Brazil, the EC-ATM model mean shows closer temporal evolution and magnitude of the MSC to the mean of the atmospheric inversions, than the EC model estimates. At a monthly time-step, the correlation of the EC-ATM with the atmospheric inversions is largely similar with the FLUXCOM RS+METEO V1 (0.99 and 0.98, respectively), only out-performing it in some tropical regions, such as BRA, EQAF, and Southwest South America (SSA). The results in Europe are very similar for the EC, EC-ATM and FLUXCOM models, with all three producing a MSC very close (RMSEs of 0.055, 0.026,

380   0.058 PgC month^{-1} respectively) to the ensemble mean of the atmospheric inversions. Overall, the EC-ATM shows a persistent improvement in the RMSE of the MSC (relative to the inversion mean) across almost all regions (Table 3).
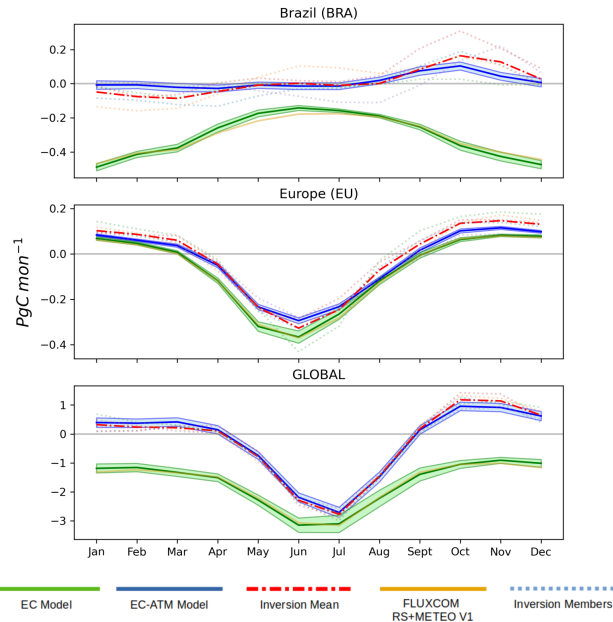
Figure 4. Mean seasonal cycle of ensemble mean of monthly NEE (PgC mon^{-1}) for a representative tropical region (Brazil BRA), extratropical region (Europe EU) and the globe, for years 2001-2017. The solid line shows the ensemble mean, and the shaded region is the mean $\pm$ the ensemble standard deviation.

4.3 Spatial variability

The spatial patterns of mean annual NEE estimated by the EC-ATM model are considerably different from those estimated by the EC model (Fig. 5). Specifically, the EC model estimates a strong mean annual sink across the tropics, while the EC-ATM model estimates more heterogeneous patterns, with sources and sinks across the tropical regions. For example, in the Amazon, the EC model estimates a strong sink of more than 1.5 TgC year^{-1} per pixel, while the mean of the atmosphere inversions shows a weak and rather homogeneous source of around 0.1 - 0.2 TgC year^{-1} per pixel, and the EC-ATM model infers a mixed pattern of strong sinks and sources. The noisy pattern in the Amazon basin estimated by the EC-ATM model may indicate some instability of the model, or an amplification of the limited signal coming from the low density of eddy-covariance sites. In tropical Africa, the EC-ATM model shows an annual source while both the EC model and the atmospheric inversions estimate a moderate to strong sink.

It should be noted that there is a known large disagreement between atmospheric inversions in the tropical regions and the location of sources and sinks varies strongly across atmospheric inversion models (Friedlingstein et al., 2022; Gaubert et al., 2019; Palmer et al., 2019)). The integration of bottom-up constraints in a unified model (EC-ATM) is not likely to resolve these spatial differences across top-down estimates, but rather to achieve NEE that is in between top-down and bottom-up approaches, as exemplified by the patterns in Figure 5.

Similar to the results of global IAV, the EC-ATM and EC model estimates display much weaker year-to-year variability in annual NEE compared to the mean of inversions (Fig. 5). While atmospheric inversions estimate IAV of about 0.5 - 0.8 TgC year^{-1} in many tropical pixels, eastern North America and western Eurasia, the EC and the EC-ATM models does not exceed 0.38 and 0.42 TgC year^{-1} respectively. This difference highlights the difficulty that both the EC and EC-ATM models have in capturing the magnitude of the interannual variability, as also shown in Jung et al. (2020).

We further evaluate the spatial distribution of NEE for four selected months representative of different seasons: January, April, July and October (Fig. 6). The distribution of the EC-ATM mean monthly flux from years 2001 to 2017 shows the reduced flux in the tropics throughout the year (red regions in the right column). The EC model estimates a stronger sink than the EC-ATM model in the tropical regions and over the whole year, but especially during the wet season (October). On the contrary, the EC-ATM model indicates a stronger sink in the Northern Hemisphere during boreal summer (North America, Europe, northern Eurasia in July), and in semi-arid regions in Southern Africa and Oceania during most of the year.

The spatial distribution of mean annual and monthly NEE estimated by the EC-ATM model is largely consistent with that estimated by the EC model, while reducing the tropical sink. EC-ATM shows some irregularities in the spatial patterns where there is insufficient information in either the eddy-covariance data or atmospheric inversions to robustly localize the NEE.
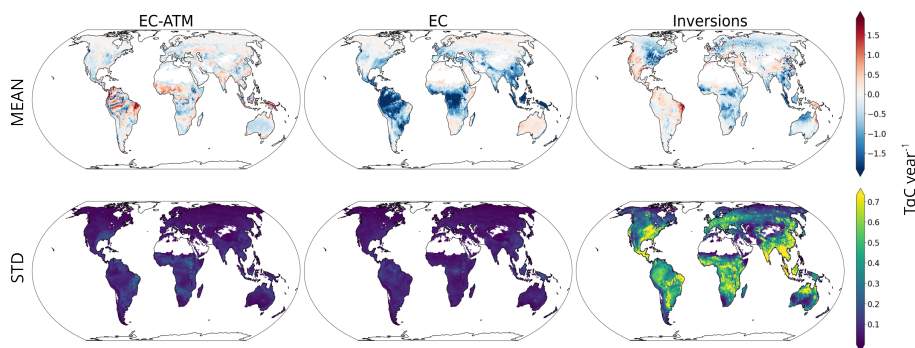
Figure 5. Mean (top panel) and standard deviation (bottom-panel) of the ensemble mean annual NEE from the EC-ATM and EC models, compared to the ensemble mean of the atmospheric inversions. The top row shows the per-pixel mean of the annual NEE. The bottom row shows the per-pixel standard deviation of the annual NEE in TgC m^{-2} year^{-1}.

We then compare the monthly pixel-level correlation between NEE estimated by the EC and EC-ATM models and by the atmospheric inversions (Fig. 7). Note that pixel-level estimates of NEE by atmospheric inversions are not expected to be robust. Nevertheless, this analysis allows to better understand how the EC-ATM model learns spatiotemporal variability in NEE. We find that both the EC-ATM and EC models correlate well with the inversion mean in the extratropics where both products are better constrained by observations, with temporal correlations greater than 0.5. In the tropics, the EC model has a negative correlation with the inversion mean in the tropics (ca. -0.4), while the EC-ATM model estimates weak but positive correlations with the inversion mean (ca. 0.2 - 0.4). While there is low confidence in atmospheric inversions' estimates at the pixel level, nevertheless, this result shows that including regionally-aggregated top-down constraints in the EC-ATM model
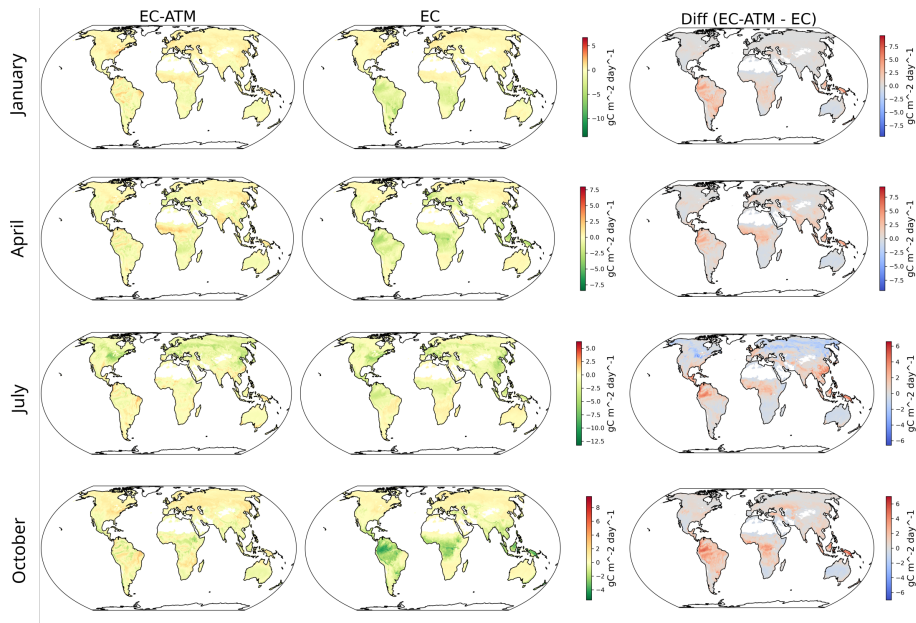
Figure 6. Monthly mean fluxes for 2001-2017 for four selected months. The left column shows the EC model results, the middle column shows EC-ATM model results and the right column shows the difference between them (EC-ATM - EC).

results in changes in the temporal variability also at pixel level (where EC-ATM pixel-level variability departs from FLUXCOM
420 RS+METEO V1, lower right panel in Figure 7). This demonstrates that the model is not simply learning a bias correction, but a new pattern of land flux that incorporates some but not all of the information in the atmospheric inversions.
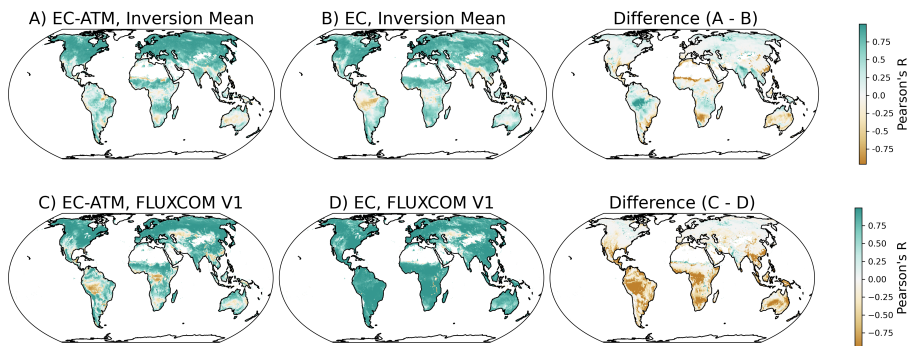
Figure 7. Spatial patterns of the per-pixel temporal correlation (Pearson's R) between EC-ATM and EC monthly NEE and the atmospheric inversion monthly mean (upper row) and FLUXCOM RS+METEO V1 monthly results (Jung et al., 2020) (bottom row), The right column shows the difference between the EC-ATM and EC correlation.

4.4 In-situ model comparison

The comparison of EC and EC-ATM modelled NEE at the eddy-covariance sites (Fig. 8) shows that the EC-ATM and EC
models have a similar RMSE with observed NEE globally. Both the EC and EC-ATM model RMSE performances of 1.349 gC
m^{-2} day^{-1} and 1.321 gC m^{-2} day^{-1} respectively are similar to the median results of RMSE for NEE using the setup in Tramontana
et al. (2016) of 1.298 gC m^{-2} day^{-1}. The optimization of the model hyperparameters (i.e. number of neurons, learning rate) for
EC-ATM performance based on the large-scale top-down constraints may lead to a slight underperfomance when these are also
applied in the EC model.

The RMSE of inferred NEE at the eddy-covariance tower level for the EC and EC-ATM models is very similar across tower
sites globally, and by the PFT that is most represented in the pixel, or majority PFT. A breakdown of tower performance by
the pixel-majority PFT is available in the appendix C2. This may indicate that, for the majority of tower sites, the information
of regional NEE by atmospheric inversions acts as a complementary constraint, improving, albeit very slightly, the EC-ATM
model estimates of NEE across the full set of eddy-covariance observations considered here.

At individual tower sites the EC-ATM model can learn different responses than the EC model. At tower BR-Ji2 (see Fig.
9 left) the EC and EC-ATM models show very similar skill in capturing variability of NEE measurements, both estimating a
smaller sink compared to the tower observations. At BR-Ma2 (see Fig. 9 right), the EC-ATM model learns an overall smaller
sink, compared with the EC model and tower observations, increasing the bias relative to NEE measurements at the tower site
compared with the EC. This demonstrates that the information from the atmospheric constraint can act as a non-complementary
constraint. Here it reduces the accuracy of the EC-ATM model at the EC tower level as it learns the smaller regional tropical
sink from the atmospheric information.
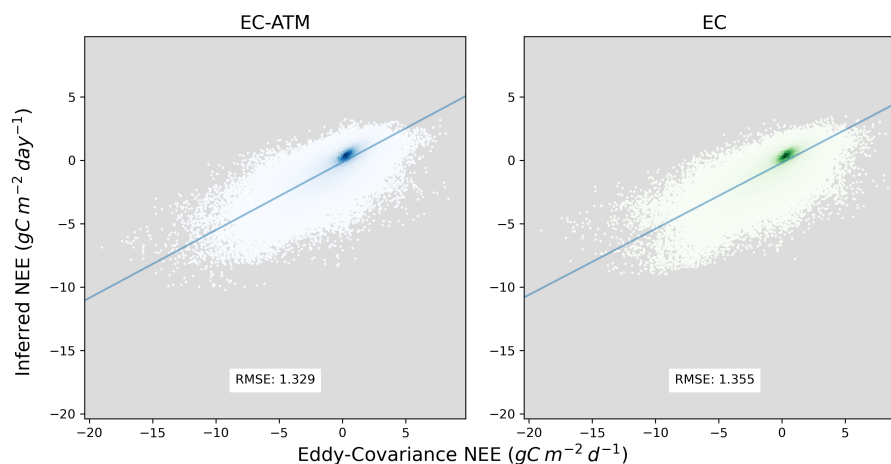
Figure 8. Comparison of inference of daily NEE from the EC-ATM and EC models with corresponding tower observations, across the whole
set of available eddy-covariance observations. The X axes show the eddy-covariance observations in gC m^{-2} day^{-1}, the Y axes show the NEE
EC-ATM and NEE EC in gC m^{-2} day^{-1}.
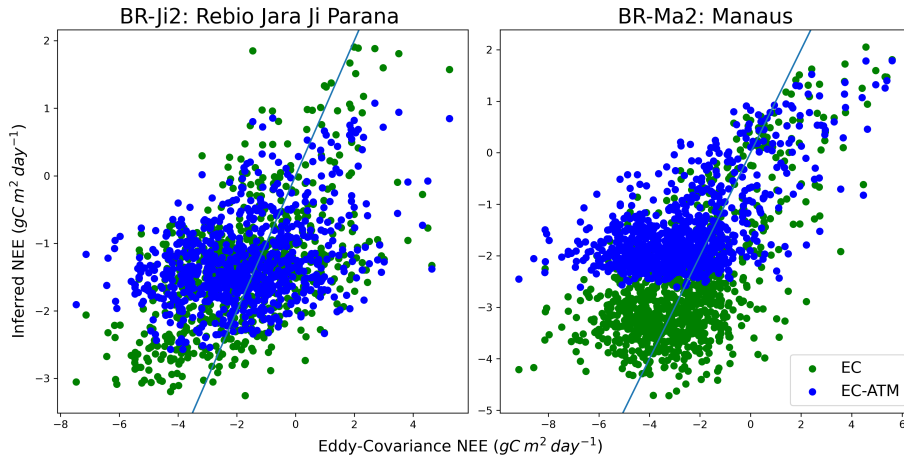
Figure 9. Comparison of EC and EC-ATM model output at two Brazilian eddy-covariance sites. The Y axes are the NEE EC-ATM and NEE EC in gC m^{-2} day^{-1}, the X axes are the eddy-covariance observations in gC m^{-2} day^{-1}. The line is the 1-to-1 line. This figure shows the different learned response in the EC-ATM model (blue) from the atmospheric constraint at a tower location compared with the EC model and the tower observations. The left-hand panel show where the learned response is similar. The right-hand panel shows where the atmospheric constraint was not complementary with the eddy-covariance constraint, and the model has a larger bias than the EC model when compared with tower observations.

5   Discussion

In this study, we aimed to evaluate the hypothesis that combining top-down and bottom-up estimates of land-atmosphere carbon fluxes could contribute to improve the estimates of regional and global land carbon sinks (Jung et al., 2020; Anav et al., 2015). To do this, we created a data-driven flux model trained on the observed NEE from eddy-covariance sites. Then, in parallel, we tested the effect of adding an additional top-down constraint based on the regional integrals of NEE from an ensemble of atmospheric inversions to the model's objective function used in NEE optimization. Our results show that adding regional atmospheric constraints to a bottom-up data-driven flux model improves NEE estimates from monthly to inter-annual and local to global scales. This approach minimizes the limitations of both top-down and bottom-up systems. It yields a model that preserves the local, small-scale view of the bottom-up approach while bringing the regional and globally integrated results inline with other best estimates of NEE.

At the global annual level the EC-ATM results show much closer correspondence with the GCPB22 residual land sink (Fig. 2). This indicates that the EC-ATM model, which runs at the pixel level with no additional larger spatial or temporal context, has learned a new response from the drivers at the ecosystem level, leading to a a more realistic global integral of NEE. This demonstrates the efficiency of the atmospheric constraint, and its ability to appropriately transmit information from the region down to the ecosystem level. At the monthly regional level, the EC-ATM outperforms the EC and FLUXCOM RS+METEO V1 results in Pearson's R and RMSE (see Table 3). The EC-ATM model improves the seasonal pattern and magnitude of the MSC across regions (see Fig. C1) when compared with other the EC model and FLUXCOM. This indicates

that the additional information does not create a simple global or regional bias-correction term, but rather a more complex
constraint that varies effectively in both space and time. This is also evident in the correction at some tower locations (Fig. 9).
460 The difference between the observed and inferred NEE by the EC and EC-ATM models at specific EC towers demonstrates
that the atmospheric information acts locally during training, and can provide both complementary and non-complementary
information to the constraint from eddy-covariance.

The impact of the complementary/non-complementary function of the combined constraints in the final model can be seen
in the differences between tropical and extratropical regions. In general, the EC, EC-ATM, and FLUXCOM results are quite
465 similar in extratropical region, especially in the northern extratropics where the eddy-covariance network is dense and the
eddy-covariance data collection method is more robust, and the atmospheric inversions have lower uncertainty (see USA, EU
in Tables 2, 3, and Figure C1). Here, the two constraints are providing complementary information, and the eddy-covariance
data appears sufficient. In tropical regions, where the eddy-covariance record is sparse, eddy-covariance measurements are
more prone to error, and the inversions are more uncertain, the EC-ATM model learns a temporally and spatially dependant
470 mixing of the two constraints, which may or may not work in a complementary way. The results for Brazil (BRA) shows that
the EC-ATM model has corrected its response at the annual and seasonal time frame, but spatially (Figs. 7, 5), we see that the
spatial and temporal distribution of NEE is different both from the inversion ensemble and FLUXCOM RS+METEO V1.

We note that this approach only partly resolved some of the weaknesses of data-driven models. The EC-ATM shows very
limited improvement relative to the EC model and the FLUXCOM RS+METEO V1 (Jung et al., 2020), in reducing the un-
475 derestimation of NEE IAV magnitude, compared with atmospheric inversions and the land sink from global carbon budgets
(Friedlingstein et al., 2022). This may be due to the fact that the optimization of the neural network and formulation of the
driver variables is the same as used in FLUXCOM RS+METEO V1. Specifically, the driver variables in both the EC and
EC-ATM models are largely based on mean season cycles of the underlying remotely-sensed data (Table A2), or with data
derived from seasonal metrics, such as the minimum and range of the mean season cycle of water availability. This limits the
480 amount of information available for the EC or EC-ATM model to learn about the IAV component of the signal. Moreover, the
magnitude of IAV is small, compared to the seasonality of NEE. Therefore, since we optimize fluxes at sub-annual time-scales
(daily for training term 1 and monthly for training term 2), the EC-ATM model may tend to optimize for the signal with larger
contribution to the objective function, i.e. the seasonal variability, rather than IAV.

Another reason for the IAV underestimation in both EC and EC-ATM models might be missing information in the training
485 set, for example due to the fact that semi-arid tropical regions, where the NEE is strongly impacted by climate variance,
and which account for a very large portion of IAV in the global land sink (Ahlström et al., 2015; Poulter et al., 2014), are
under-represented in the FLUXNET network. We expect the regional top-down constraint by atmospheric inversions to partly
improve this, but atmospheric inversions tend to assign less IAV to the tropical semi-arid regions than land-surface models
used in global carbon budgets (Piao et al., 2020), and also tend to infer lower sensitivity of NEE IAV to tropical water and
490 temperature variability (Wang et al., 2022).

While model structure could contribute to this issues, experiments (not shown) with varying architecture (number of neurons,
number of layers, layer connectivity), activation functions, and loss terms of the EC and EC-ATM models were not successful

in reducing the bias in the magnitude of NEE IAV. We also conducted sensitivity experiments on the objective function both structurally, and in its formulation. Structurally, the number of regions included in the atmospheric term at each step, and the number of months included at each step were varied. The best results, as indicated in the methods, was considering a full year for each region at each training step.

5.1 Uncertainty in top-down constraints

The sparsity of observations does not only concern the FLUXNET network, but also the network of atmospheric monitoring sites used by atmospheric inversions. Given the small number of tropical observations considered in top-down constraints, and the inherited uncertainty from priors and transport models, which are not directly managed by the system in this study (Baker et al., 2006), the EC-ATM model result might still be prone to large uncertainties in the tropics. Here, we discuss the handling of uncertainty in the EC-ATM model in more detail.

The system presented in this study is still dependent on the priors and transport models in the underlying atmospheric inversions, and is still subject to the underlying uncertainty of these data, particularly in the tropics where both bottom-up and top-down systems lack sufficient observations. During training, the EC-ATM model tries to account for uncertainties in two ways: the error between the inference and the inversion data is normalized in the loss 2 term by the spread of the inversion estimates by time and region, which reflects uncertainty across inversion models (Eq. 10); additionally, the model learns to weight the objective term relative to the estimated uncertainty in the atmospheric loss term, which should tend to reduce the weight where there is larger systematic disagreement between inversion systems (Eq. 12).

We perform a "one-against-many" sensitivity analysis, where we trained several models with the atmospheric constraint coming from either one inversion (zero spread), two randomly selected inversions, or three randomly selected inversions. This analysis, shown in Supplement (Figs. C3, C4, C5) allows us to evaluate how the specific inversion NEE estimates interact with the loss mixing scheme in the model training. In extra-tropical regions where the eddy-covariance and atmospheric observations are dense, the specific inversion NEE trained on does not critically influence the model response. There is strong correspondence between all EC-ATM instances with individual or smaller groups of inversions and the full inversion ensemble mean. In tropical regions, there is definite movement towards the specific inversion NEE trained on (the dotted lines in the Figs. C3, C4, C4). This response is balanced against the model initialization and the learned weighting scheme during training. At the global level there is a closer agreement between the trained models and the target inversions, as the relative noise of the tropical regions is dampened by the more consistent extratropical regions. As additional inversions are added, there is a tendency for EC-ATM NEE to move closer to the inversion mean, as the target for optimization contains a larger amount of NEE values from the ensemble of inversions. This analysis demonstrates that the EC-ATM model inherits the uncertainty from the ensemble of atmospheric inversions, with largest uncertainty remaining in the tropical regions where the available observations for both top-down and bottom-up is lacking.

Our results also show the potential for a confounding effect from the training process. The EC-ATM model is a learned statistical response between the drivers and the training data. There are mismatches between the EC-ATM inference of NEE and the atmospheric data used for the top-down constraint. The atmospheric inversion NEE data, although adjusted for fossil

fuel, fire and riverine fluxes, still implicitly includes disturbance and trade fluxes, along with other flux components that are not seen by eddy-covariance measurements, and are not accounted for in our model. This means that in reducing the loss terms (Eq 14), these flux components are implicitly incorporated into the EC-ATM inference, although the model lacks the necessary process information which is not included in the drivers. Furthermore, given the statistical nature of the network used and the training process, the data-driven model should not be considered an analog for a process-based model, where individual terms can be more easily backed out. Training using our double constraint should become less confounded as more additional spatially-explicit flux components become available. However, despite these data mismatches, training a data-driven flux model using a dual constraint does create a useful estimate of the NEE at multiple scales.

6 Outlook and conclusions

Our study aimed to demonstrate that adding an atmospheric top-down constraint can positively impact the evolution of a bottom-up data-driven flux model during training, leading to meaningful improvement in local to global NEE estimates. The study demonstrated the positive impact of regional atmospheric information on the training of a well-established data-driven flux model (Jung et al., 2020), and demonstrated the applicability of using a observational constraints of NEE at different spatial scales.

In this study, we combined regional integrals of NEE from atmospheric inversions with in-situ NEE from eddy-covariance measurements. We note, however, that these could be multiple data streams at different scales (temporal, spatial), or in different formats (grid, point). Incorporating these different data streams would require different model formulations, potentially including neural network architecture and objective functions, as well as data-driven, or physics-based bridge models to create the link from the data-driven flux-model to these new data.

This multi-scale approach proposed here may allow us to leverage large volumes of additional data for constraining a data-driven flux model. By substituting a physical model for the statistical bridge model used here, a double-constraint data-driven flux model could generate an inference of NEE across diverse temporal and spatial scales. Because, unlike the statistical bridge models, these additional data could vary with the local meteorology, covering a range of biomes, the data-driven flux model would see a more diverse training set. This could improve the performance of the data-driven flux model by learning from a more representative distribution of the driver variables across the land surface. In the future, this logic could be used for a variety of datasets, for example by pairing the archive of eddy-covariance observations with tall-tower observations of the mole fraction of CO_2, or with novel 'flux towers in the sky' (Schimel et al., 2019) estimates from satellite retrievals of total column CO_2 (XCO_2).

Data availability. Inversion data: https://doi.org/10.18160/7AH8-K1X4

Table A1. RECCAP2 region ID, names and abbreviations

ID	Region	Abbreviation
0	United States	USA
1	Canada	CAN
2	Central America	CAM
3	Northern South America	NSA
4	Brazil	BRA
5	Southwest South America	SSA
6	Europe	EU
7	Northern Africa	NAF
8	Equatorial Africa	EQAF
9	Southern Africa	SAF
10	Russia	RUS
11	Central Asia	CAS
12	Mideast	MIDE
13	China	CHN
14	Korea and Japan	KAJ
15	South Asia	SAS
16	Southeast Asia	SEAS
17	Oceania	OCE

Appendix A: Data
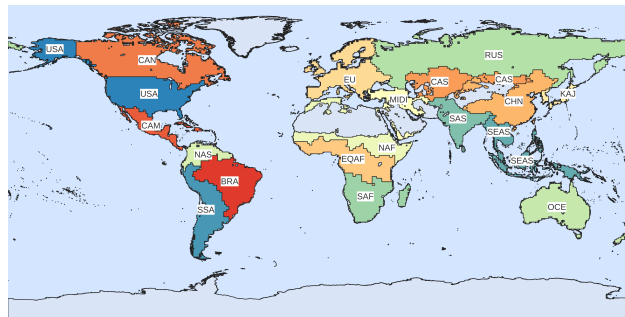
Figure A1. RECCAP2 Regions

Table A2. Driver variables used for the data-driven EC and EC-ATM models and the calculation of the drivers from the base variables above. The global dataset uses only the MODIS and ERA5 data, while the data used at the eddy-covariance sites also uses meteorological observations from the tower instruments. See Tramontana et al. (2016) for a full discussion.

Name	Variable	MSC calculation period	(Source)
WAI2	Water Availability Index		Calculated from a water balance model derived from MODIS variables (see Tramontana et al. (2016) supplement S3 for full model description)
MSC_EVIRpot	Mean Season Cycle(EVI $\times$ R_g)	2001–2012	MODIS (EVI), ERA5 (R_g)
MSC_FparLST	Mean Season Cycle(fAPAR $\times$ LST_{day})	2001–2012	MODIS
MIN_MSC_NDWI	Min(Mean Season Cycle(NDWI))	2001–2012	MODIS
AMP_Band4	Amplitude(band 4 reflectance)		MODIS
MSC_LST_Night	Mean Season Cycle(LST_{night})	2001–2012	MODIS
Rg_VIMSC	Mean Season Cycle(NDVI) $\times$ R_{pot}	2001–2012	MODIS (NDVI), ERA5 (R_g)
AMP_MSC_NDVI	Amplitude(Mean Season Cycle(NDVI))	2001–2012	MODIS
Tair	Air Temperature	2001–2012	ERA5
AMP_MSC_WAI	Amplitude(Mean Season Cycle(WAI))	2001–2012	ERA5

All variables: Daily values 2000-2017, 0.5°spatial resolution

Table B1. Hyperparameters for reported EC and EC-ATM model runs

Input shape	10
Latent dimension	32
Learning rate	0.003
EC batch	10000

Appendix B: Technical Implementation
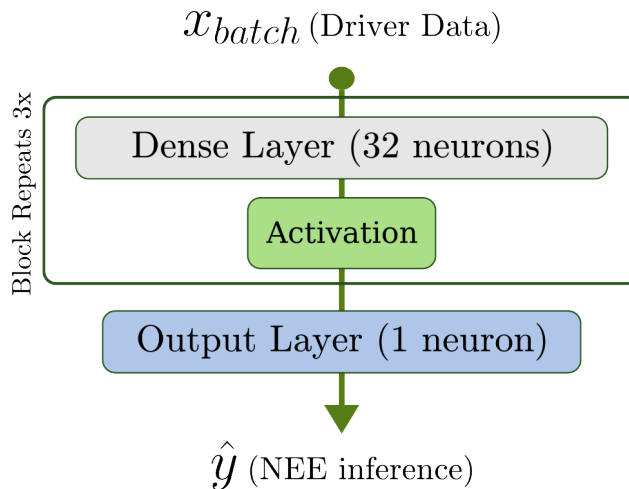
Figure B1. Model Architecture, the model is a feed-forward neural network, or a set of fully connected network layers. The fully-connected layers consist of nodes or 'neurons', which are exposed to the output of all neurons in the previous layer. Non-linearity is introduced by passing each node output through a non-linear activation function. Our network is a set of three fully-connected layers with the ReLU activation function (Agarap, 2019).
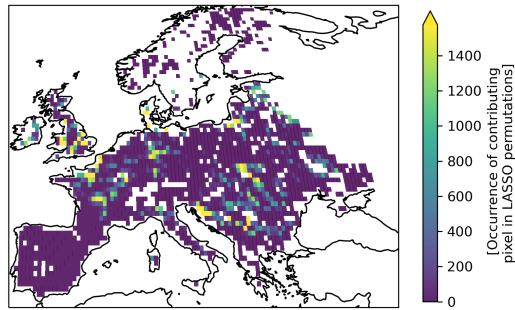
Figure B2. Robustness of contribution pixel selection: A Heat map of pixel inclusion in the sparse linear model using Lasso regression. Values represent the log-scaled number of pixel inclusions in the non-zero set of parameters across 500 regressions using a randomized subset of the data. Pixels that are most often included provide a more important constraint to the calculation of a regionally summed NEE, minimizing Eq. 1
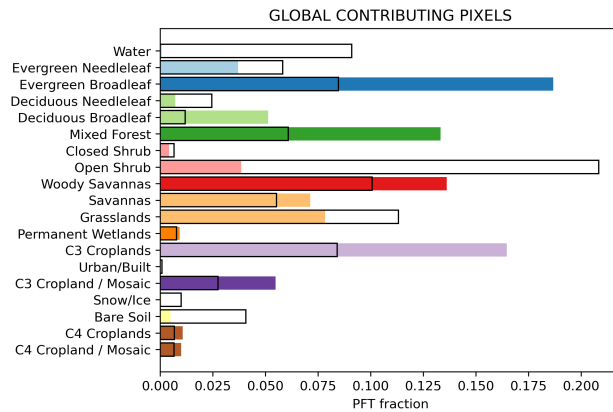
Figure B3. The representation of PFTs across all contributing pixels in all regions. All PFTs are the majority type per pixel. This image shows the relative number of times a certain PFT is included in the optimal set of contributing pixels which construct a regional integral of NEE, when selecting from all global land pixels. The black outlines show the proportion of that majority PFT type globally. A per-region analysis of PFT inclusion is available in appendix B4
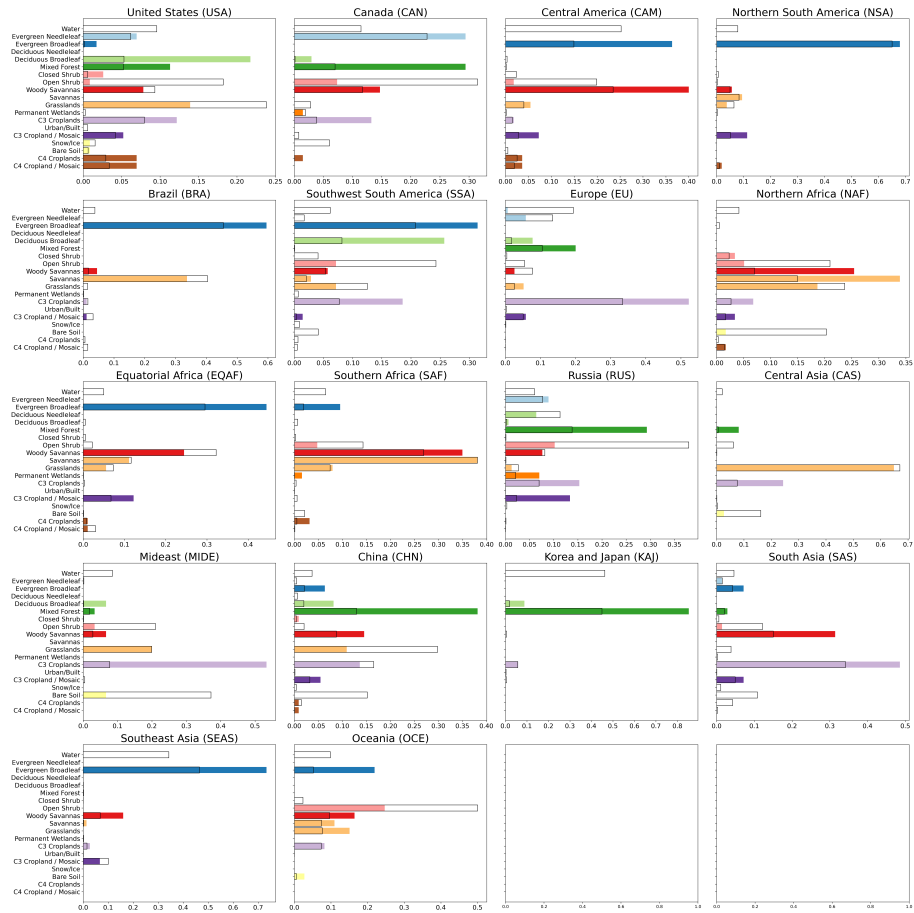
Figure B4. Regional composition of PFT in contributing pixels

Appendix C: Results

Table C1. Estimates of NEE from land from recent studies

Study	Land sink mean (Pg C yr-1)	IAV magnitude	Period
Crisp et al. (2022), 2022	-2.2	0.6	2000-2009
Friedlingstein et al. (2022)	-2.91	0.81	2000-2017
Gaubert et al. (2019)	Northern ExtraTropics: -2.17 Tropics+Southern ExtraTropics: -0.06	Northern ExtraTropics: 0.36 Tropics+Southern ExtraTropics: 0.11	2004–2014
Ruehr et al. (2023)	-3.1	0.6	2010-2019
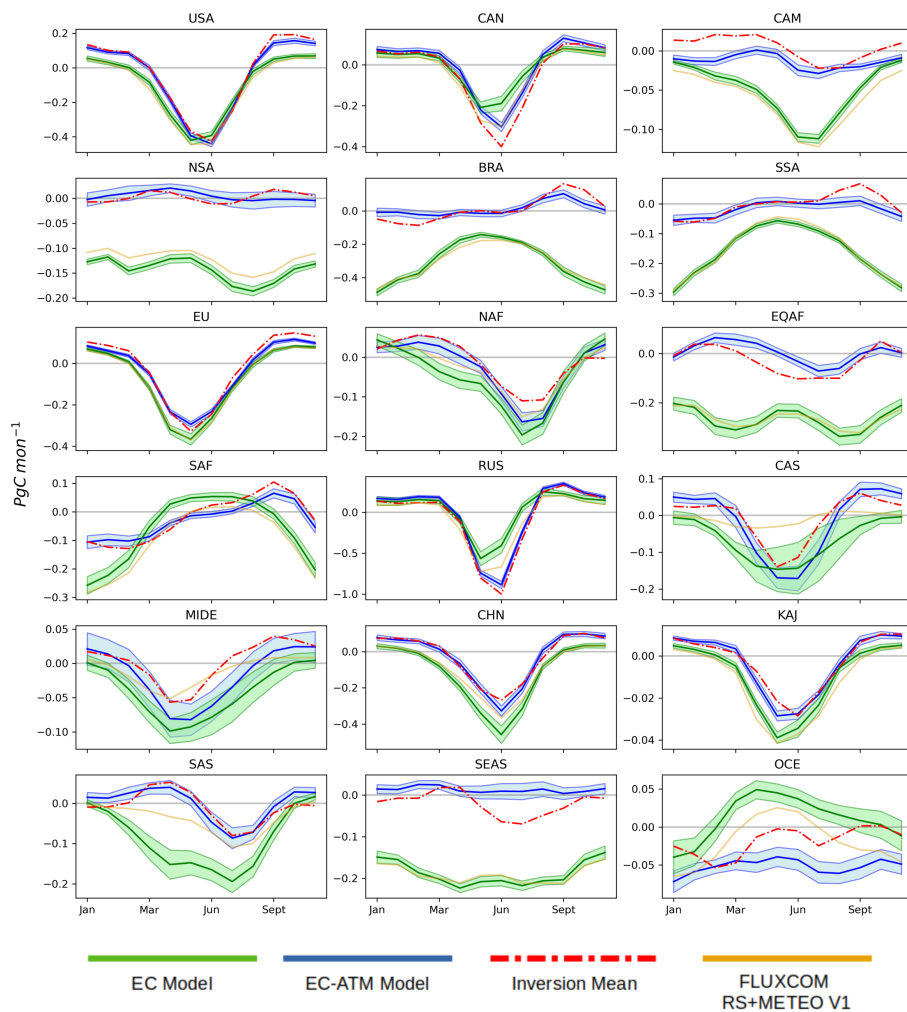
Figure C1. MSC of the ensemble mean of all regions in PgC mon^{-1}. The solid line is the ensemble mean, and the shaded region is the mean $\pm$ the ensemble standard deviation.
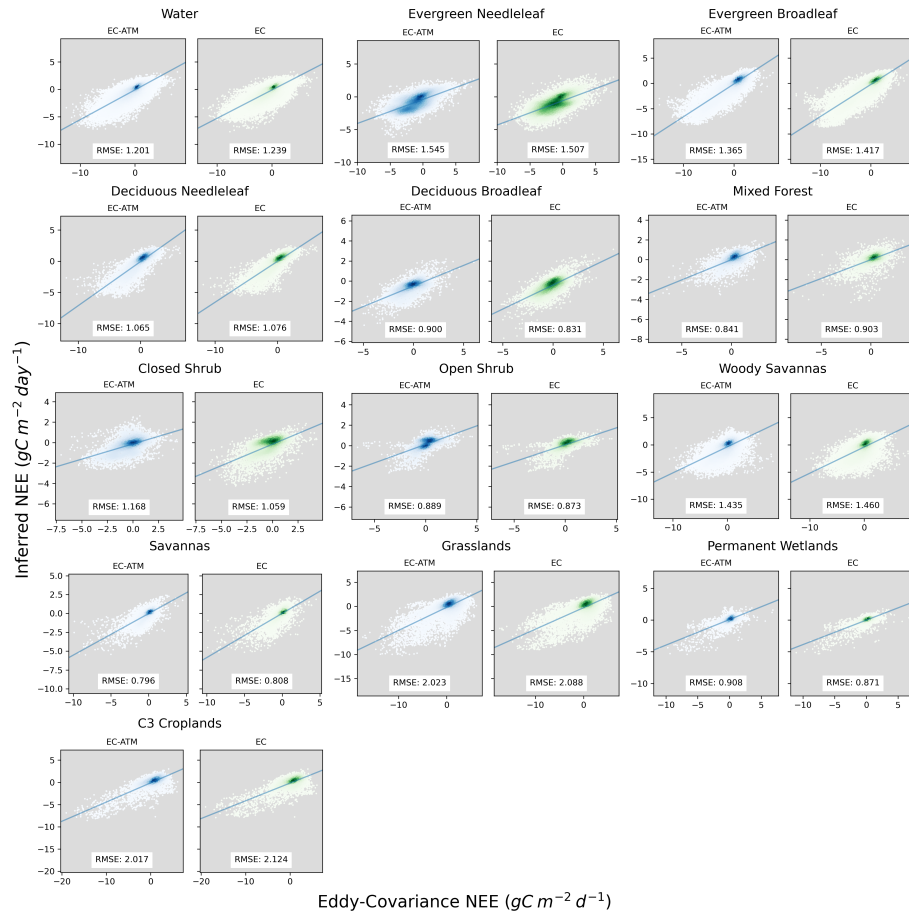
Figure C2. Scatter-plots of eddy-covariance NEE (x axis) and inferred NEE (y axis) by PFT and model (ATM-EC, EC).
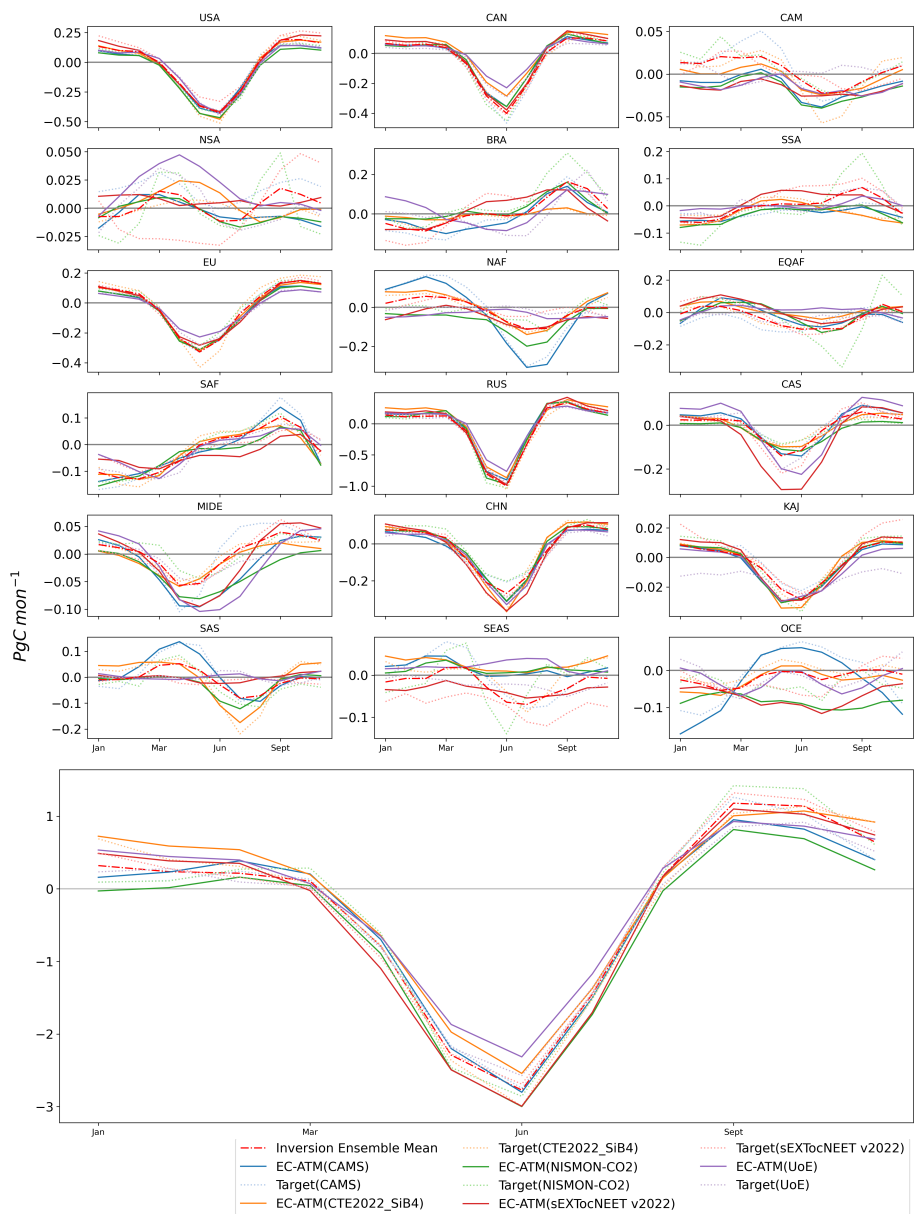
Figure C3. The MSC results for the 'one-against-many' training runs. A separate EC-ATM model was trained for each individual inversion system to test the impact of the full ensemble and loss normalization in the full study. The solid lines are the EC-ATM models trained using the named inversion system, and the dotted lines are MSC of the inversion system regional and global integrals. For all panels, the x-axis is the yearly cycle, and the y axis is PgC mon^{-1}.
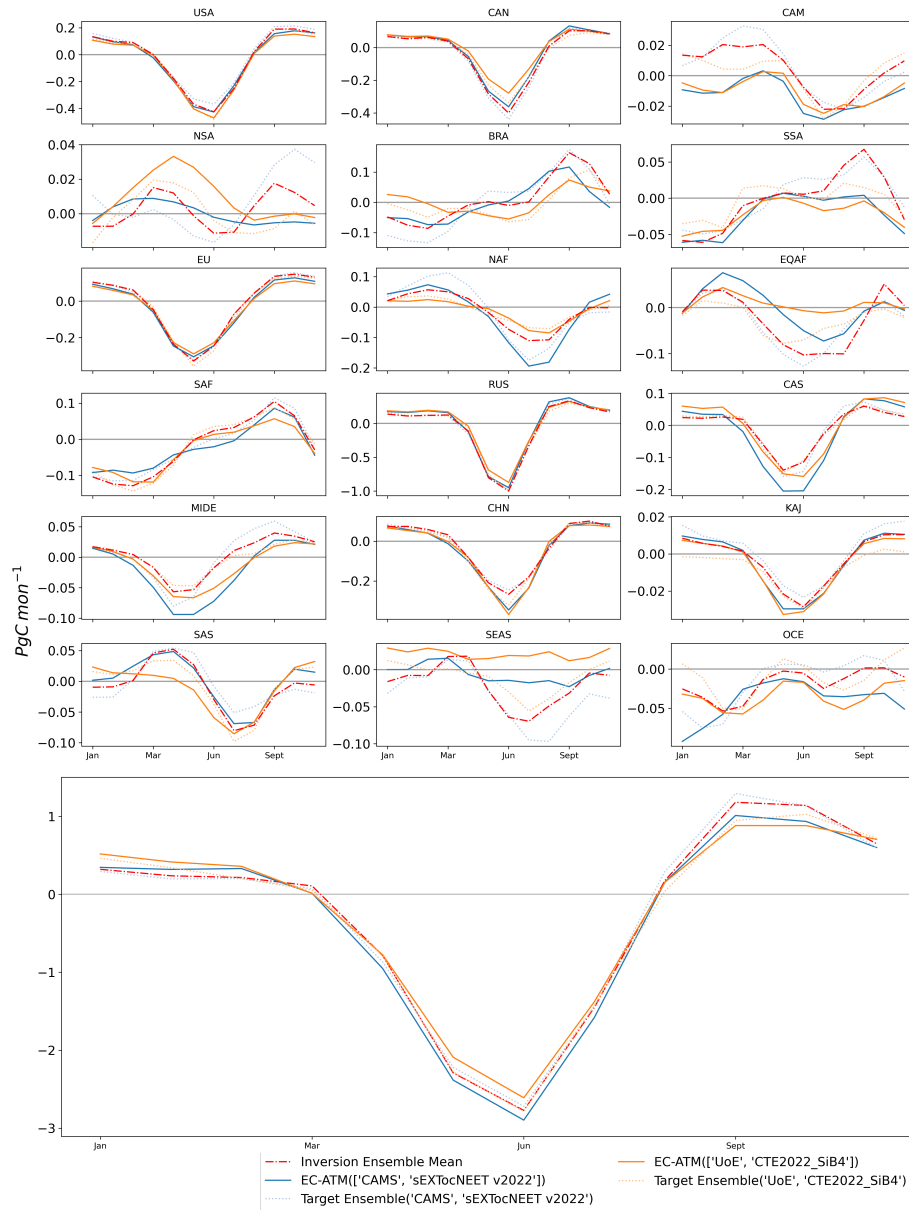
Figure C4. The MSC results for the 'one-against-many' training runs, with EC-ATM models optimized against two inversion systems. The solid lines are the EC-ATM models trained using the named pair of inversion systems, and the dotted lines are MSC of the named pair's mean regional and global integrals. For all panels, the x-axis is the yearly cycle, and the y axis is PgC mon^{-1}.
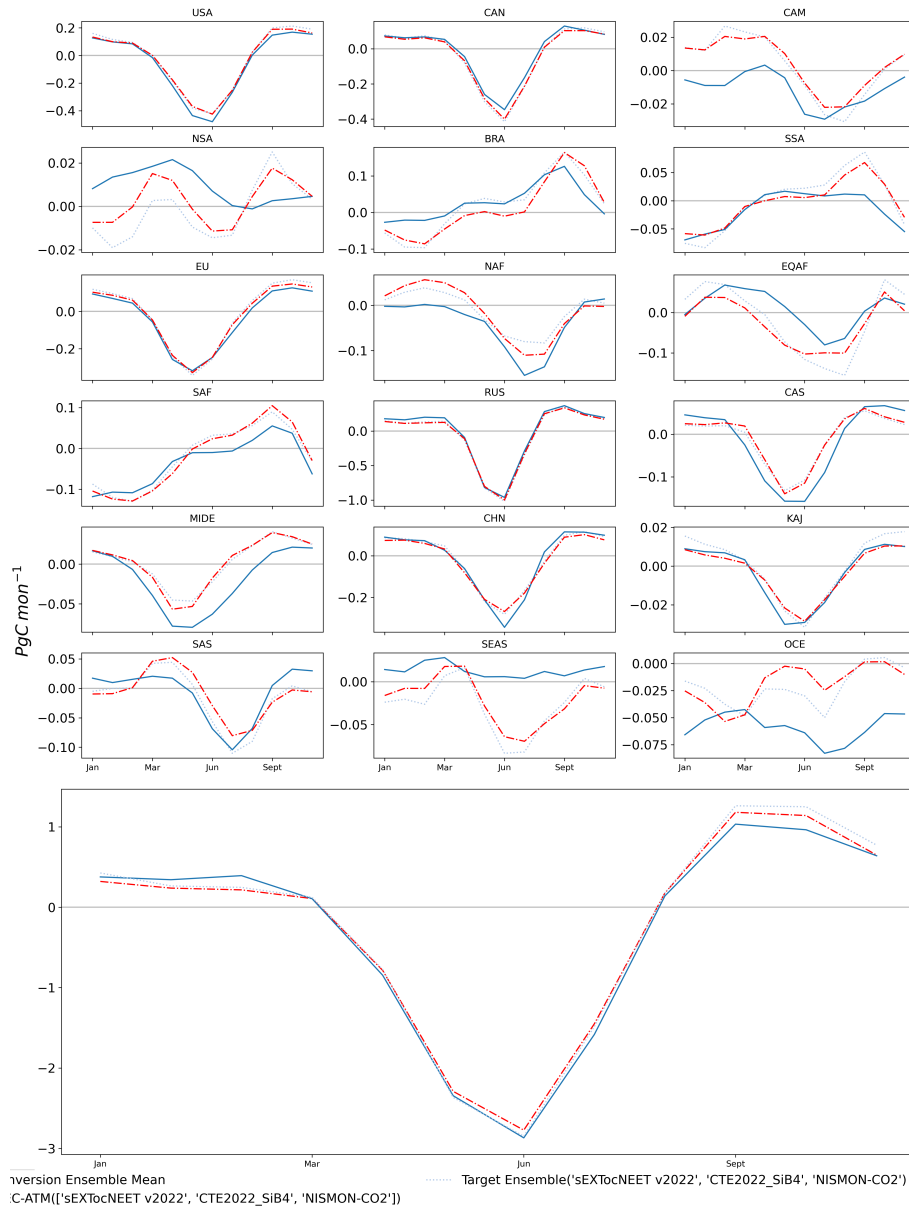
Figure C5. The MSC results for the 'one-against-many' training runs, with an EC-ATM model optimized against three inversion systems. The solid line is the EC-ATM model trained using the named set of inversion systems, and the dotted line is the MSC of the named set's mean regional and global integrals. For all panels, the x-axis is the yearly cycle, and the y axis is PgC mon^{-1}.

Author contributions. SU, AB, MR, FG designed the study. SU performed the analysis, and drafted the manuscript. AB, MR, WP provided analysis and support. BK provided expertise with the machine learning framework. All authors revised and edited the text.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We would like to thank Martin Jung, Jakob A. Nelson, Sophia Walther, and the FLUXCOM team for their structural support, feedback and discussion. The Authors would like to thank the producers of the Inversion data included in this study: Ingrid Luijkx and Wouter Peters (CTE), Frederic Chevallier and the Copernicus Atmosphere Monitoring Service (CAMS), Christian Roedenbeck (Jena Carboscope sEXTocNEET), Yosuke Niwa (NISMON-CO2), and Liang Feng and Paul Palmer (UoE). We would like to thank the team behind the European Research Council (ERC) Synergy Grant 'Understanding and modeling the Earth System with Machine Learning (USMILE)' for their feedback and discussion.

This work used eddy covariance data acquired by the FLUXNET community and in particular by the following networks: AmeriFlux (U.S. Department of Energy, Biological and Environmental Research, Terrestrial Carbon Program (DE-FG02-04ER63917 and DE-FG02-04ER63911)), AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada (supported by CFCAS, NSERC, BIOCAP, Environment Canada, and NRCan), GreenGrass, KoFlux, LBA, NECC, OzFlux, TCOS-Siberia, USCCC. We acknowledge the financial support to the eddy covariance data harmonization provided by CarboEuropeIP, FAO-GTOS-TCO, iLEAPS, Max Planck Institute for Biogeochemistry, National Science Foundation, University of Tuscia, Université Laval and Environment Canada and US Department of Energy and the database development and technical support from Berkeley Water Center, Lawrence Berkeley National Laboratory, Microsoft Research eScience, Oak Ridge National Laboratory, University of California - Berkeley, University of Virginia

References

Agarap, A. F.: Deep Learning using Rectified Linear Units (ReLU), https://doi.org/10.48550/arXiv.1803.08375, arXiv:1803.08375 [cs, stat], 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, _eprint: https://www.science.org/doi/pdf/10.1126/science.aaa1668, 2015.

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., Murray-Tortarolo, G., Papale, D., Parazoo, N. C., Peylin, P., Piao, S., Sitch, S., Viovy, N., Wiltshire, A., and Zhao, M.: Spatiotemporal patterns of terrestrial gross primary production: A review, Reviews of Geophysics, 53, 785–818, https://doi.org/10.1002/2015RG000483, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2015RG000483, 2015.

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 inter-comparison: Impact of transport model errors on the interannual variability of regional CO2 fluxes, 1988–2003, Global Biogeochemical Cycles, 20, https://doi.org/10.1029/2004GB002439, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2004GB002439, 2006.

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.

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, Journal of Geophysical Research: Atmospheres, 110, https://doi.org/10.1029/2005JD006390, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2005JD006390, 2005.

Chu, H., Baldocchi, D. D., John, R., Wolf, S., and Reichstein, M.: Fluxes all of the time? A primer on the temporal representative-ness of FLUXNET, Journal of Geophysical Research: Biogeosciences, 122, 289–307, https://doi.org/10.1002/2016JG003576, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2016JG003576, 2017.

Ciais, P., Rayner, P., Chevallier, F., Bousquet, P., Logan, M., Peylin, P., and Ramonet, M.: Atmospheric inversions for estimating CO2 fluxes: methods and perspectives, Climatic Change, 103, 69–92, https://doi.org/10.1007/s10584-010-9909-3, 2010.

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), Geoscientific Model Development, 15, 1289–1316, https://doi.org/10.5194/gmd-15-1289-2022, publisher: Copernicus GmbH, 2022.

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?, Reviews of Geophysics, 60, e2021RG000 736, https://doi.org/10.1029/2021RG000736, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021RG000736, 2022.

Deng, Z., Ciais, P., Tzompa-Sosa, Z. A., Saunois, M., Qiu, C., Tan, C., Sun, T., Ke, P., Cui, Y., Tanaka, K., et al.: Comparing national greenhouse gas budgets reported in UNFCCC inventories against atmospheric inversions, Earth System Science Data Discussions, 2021, 1–59, 2021.

Di Giuseppe, F., Rémy, S., Pappenberger, F., and Wetterhall, F.: Using the Fire Weather Index (FWI) to improve the estimation of fire emissions from fire radiative power (FRP) observations, Atmospheric Chemistry and Physics, 18, 5359–5370, https://doi.org/10.5194/acp-18-5359-2018, publisher: Copernicus GmbH, 2018.

Feng, L., Palmer, P. I., Parker, R. J., Deutscher, N. M., Feist, D. G., Kivi, R., Morino, I., and Sussmann, R.: Estimates of European uptake of CO_2 inferred from GOSAT X_{CO_2} retrievals: sensitivity to measurement bias inside and outside Europe, Atmospheric Chemistry and Physics, 16, 1289–1302, https://doi.org/10.5194/acp-16-1289-2016, publisher: Copernicus GmbH, 2016.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X.: MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets, Remote Sensing of Environment, 114, 168–182, https://doi.org/10.1016/j.rse.2009.08.016, 2010.

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Bakker, D. C. E., Hauck, J., Le Quéré, C., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., Bates, N. R., Becker, M., Bellouin, N., Bopp, L., Chau, T. T. T., Chevallier, F., Chini, L. P., Cronin, M., Currie, K. I., Decharme, B., Djeutchouang, L. M., Dou, X., Evans, W., Feely, R. A., Feng, L., Gasser, T., Gilfillan, D., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, O., Harris, I., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Luijkx, I. T., Jain, A., Jones, S. D., Kato, E., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lienert, S., Liu, J., Marland, G., McGuire, P. C., Melton, J. R., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Niwa, Y., Ono, T., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Rosan, T. M., Schwinger, J., Schwingshackl, C., Séférian, R., Sutton, A. J., Sweeney, C., Tanhua, T., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F., van der Werf, G. R., Vuichard, N., Wada, C., Wanninkhof, R., Watson, A. J., Willis, D., Wiltshire, A. J., Yuan, W., Yue, C., Yue, X., Zaehle, S., and Zeng, J.: Global Carbon Budget 2021, Earth System Science Data, 14, 1917–2005, https://doi.org/10.5194/essd-14-1917-2022, publisher: Copernicus GmbH, 2022.

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, Agricultural and Forest Meteorology, 263, 292–307, https://doi.org/10.1016/j.agrformet.2018.09.001, 2018.

Gaubert, B., Stephens, B. B., Basu, S., Chevallier, F., Deng, F., Kort, E. A., Patra, P. K., Peters, W., Rödenbeck, C., Saeki, T., Schimel, D., Van der Laan-Luijkx, I., Wofsy, S., and Yin, Y.: Global atmospheric CO2 inverse models converging on neutral tropical land exchange, but disagreeing on fossil fuel and atmospheric growth rate, Biogeosciences, 16, 117–134, https://doi.org/10.5194/bg-16-117-2019, publisher: Copernicus GmbH, 2019.

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, Agricultural and Forest Meteorology, 250-251, 90–101, https://doi.org/10.1016/j.agrformet.2017.12.186, 2018.

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, Scientific Data, 6, 74, https://doi.org/10.1038/s41597-019-0076-8, bandiera_abtest: a Cc_license_type: cc_publicdomain Cg_type: Nature Research Journals Number: 1 Primary_atype: Re-

search Publisher: Nature Publishing Group Subject_term: Climate sciences;Environmental sciences;Hydrology Subject_term_id: climate-sciences;environmental-sciences;hydrology, 2019.

650 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, publisher: Copernicus GmbH, 2020.

655 Kaminski, T. and Heimann, M.: Inverse Modeling of Atmospheric Carbon Dioxide Fluxes, Science, https://doi.org/10.1126/science.294.5541.259a, publisher: American Association for the Advancement of Science, 2001.

Keeling, C. D., Bacastow, R. B., Carter, A. F., Piper, S. C., Whorf, T. P., Heimann, M., Mook, W. G., and Roeloffzen, H.: A three-dimensional model of atmospheric CO2 transport based on observed winds: 1. Analysis of observational data, in: Aspects of Climate Variability in the Pacific and the Western Americas, pp. 165–236, American Geophysical Union (AGU), https://doi.org/10.1029/GM055p0165, _eprint:
660 https://onlinelibrary.wiley.com/doi/pdf/10.1029/GM055p0165, 1989.

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: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7482–7491, 2018.

Kondo, M., Ichii, K., Takagi, H., and Sasakawa, M.: Comparison of the data-driven top-down and bottom-up global terrestrial CO 2 ex-
665 changes: GOSAT CO 2 inversion and empirical eddy flux upscaling: Data-driven Terrestrial CO 2 Exchanges, Journal of Geophysical Research: Biogeosciences, 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-
670 down and bottom-up approaches for terrestrial CO2 budget, Global Change Biology, 26, 1068–1084, https://doi.org/10.1111/gcb.14917, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.14917, 2020.

Niwa, Y., Ishijima, K., Ito, A., and Iida, Y.: Toward a long-term atmospheric CO2 inversion for elucidating natural carbon fluxes: technical notes of NISMON-CO2 v2021.1, Progress in Earth and Planetary Science, 9, 42, https://doi.org/10.1186/s40645-022-00502-6, 2022.

Palmer, P. I., Feng, L., Baker, D., Chevallier, F., Bösch, H., and Somkuti, P.: Net carbon emissions from African biosphere dominate pan-
675 tropical atmospheric CO2 signal, Nature Communications, 10, 3344, https://doi.org/10.1038/s41467-019-11097-w, number: 1 Publisher: Nature Publishing Group, 2019.

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, publisher: Copernicus GmbH, 2013.

680 Piao, S., Wang, X., Wang, K., Li, X., Bastos, A., Canadell, J. G., Ciais, P., Friedlingstein, P., and Sitch, S.: Interannual variation of terrestrial carbon cycle: Issues and perspectives, Global Change Biology, 26, 300–318, https://doi.org/10.1111/gcb.14884, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/gcb.14884, 2020.

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,
685 600–603, https://doi.org/10.1038/nature13376, number: 7502 Publisher: Nature Publishing Group, 2014.

Ruehr, S., Keenan, T. F., Williams, C., Zhou, Y., Lu, X., Bastos, A., Canadell, J. G., Prentice, I. C., Sitch, S., and Terrer, C.: Evidence and attribution of the enhanced land carbon sink, Nature Reviews Earth & Environment, pp. 1–17, https://doi.org/10.1038/s43017-023-00456-3, publisher: Nature Publishing Group, 2023.

Rödenbeck, C., Houweling, S., Gloor, M., and Heimann, M.: CO2 flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport, Atmospheric Chemistry and Physics, 3, 1919–1964, https://doi.org/10.5194/acp-3-1919-2003, publisher: Copernicus GmbH, 2003.

Rödenbeck, C., Zaehle, S., Keeling, R., and Heimann, M.: How does the terrestrial carbon exchange respond to inter-annual climatic variations? A quantification based on atmospheric CO_2 data, Biogeosciences, 15, 2481–2498, https://doi.org/10.5194/bg-15-2481-2018, publisher: Copernicus GmbH, 2018.

Saeki, T. and Patra, P. K.: Implications of overestimated anthropogenic CO2 emissions on East Asian and global land CO2 flux inversion, Geoscience Letters, 4, 9, https://doi.org/10.1186/s40562-017-0074-7, 2017.

Schimel, D., Schneider, F. D., and JPL Carbon and Ecosystem Participants: Flux towers in the sky: global ecology from space, New Phytologist, 224, 570–584, https://doi.org/10.1111/nph.15934, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/nph.15934, 2019.

Thompson, R. L., Patra, P. K., Chevallier, F., Maksyutov, S., Law, R. M., Ziehn, T., van der Laan-Luijkx, I. T., Peters, W., Ganshin, A., Zhuravlev, R., Maki, T., Nakamura, T., Shirai, T., Ishizawa, M., Saeki, T., Machida, T., Poulter, B., Canadell, J. G., and Ciais, P.: Top–down assessment of the Asian carbon budget since the mid 1990s, Nature Communications, 7, 10 724, https://doi.org/10.1038/ncomms10724, number: 1 Publisher: Nature Publishing Group, 2016.

Tian, H., Yang, J., Lu, C., Xu, R., Canadell, J. G., Jackson, R. B., Arneth, A., Chang, J., Chen, G., Ciais, P., Gerber, S., Ito, A., Huang, Y., Joos, F., Lienert, S., Messina, P., Olin, S., Pan, S., Peng, C., Saikawa, E., Thompson, R. L., Vuichard, N., Winiwarter, W., Zaehle, S., Zhang, B., Zhang, K., and Zhu, Q.: The Global N2O Model Intercomparison Project, Bulletin of the American Meteorological Society, 99, 1231–1251, https://doi.org/10.1175/BAMS-D-17-0212.1, publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society, 2018.

Tibshirani, R.: Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society: Series B (Methodological), 58, 267–288, 1996.

Tormene, P., Giorgino, T., Quaglini, S., and Stefanelli, M.: Matching Incomplete Time Series with Dynamic Time Warping: An Algorithm and an Application to Post-Stroke Rehabilitation, Artificial Intelligence in Medicine, 45, 11–34, https://doi.org/10.1016/j.artmed.2008.11.007, 2008.

Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., and Papale, D.: Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, Biogeosciences, 13, 4291–4313, https://doi.org/10.5194/bg-13-4291-2016, publisher: Copernicus GmbH, 2016.

Valentini, R., Arneth, A., Bombelli, A., Castaldi, S., Cazzolla Gatti, R., Chevallier, F., Ciais, P., Grieco, E., Hartmann, J., Henry, M., Houghton, R. A., Jung, M., Kutsch, W. L., Malhi, Y., Mayorga, E., Merbold, L., Murray-Tortarolo, G., Papale, D., Peylin, P., Poulter, B., Raymond, P. A., Santini, M., Sitch, S., Vaglio Laurin, G., van der Werf, G. R., Williams, C. A., and Scholes, R. J.: A full greenhouse gases budget of Africa: synthesis, uncertainties, and vulnerabilities, Biogeosciences, 11, 381–407, https://doi.org/10.5194/bg-11-381-2014, publisher: Copernicus GmbH, 2014.

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, Geoscientific Model Development, 10, 2785–2800, https://doi.org/10.5194/gmd-10-2785-2017,
publisher: Copernicus GmbH, 2017.

Wang, K., Bastos, A., Ciais, P., Wang, X., Rödenbeck, C., Gentine, P., Chevallier, F., Humphrey, V. W., Huntingford, C., O'Sullivan, M., Seneviratne, S. I., Sitch, S., and Piao, S.: Regional and seasonal partitioning of water and temperature controls on global land carbon uptake variability, Nature Communications, 13, 3469, https://doi.org/10.1038/s41467-022-31175-w, number: 1 Publisher: Nature Publishing Group, 2022.

725