CEDAR-GPP: spatiotemporally upscaled estimates of gross primary productivity incorporating CO_2 fertilization

Yanghui Kang[1,2,3], Maoya Bassiouni[1,2], Max Gaber[1,4], Xinchen Lu[1,2], Trevor F. Keenan[1,2]

[1] Department of Environmental Science, Policy, and Management, University of California, Berkeley, Berkely, CA 94720, USA

[2] Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

[3] Department of Biological Systems Engineering, Virginia Tech, Blacksburg, VA 24061, USA

[4] Department of Geosciences and Natural Resource Management, University of Copenhagen, Copenhagen, 1350, Denmark.

Correspondence: Yanghui Kang (yanghuikang@vt.edu)
Trevor Keenan (trevorkeenan@berkeley.edu)

Abstract: Gross primary productivity (GPP) is the largest carbon flux in the Earth system, playing a crucial role in removing atmospheric carbon dioxide and providing carbohydrates needed for ecosystem metabolism. Despite the importance of GPP, however, existing estimates present significant uncertainties and discrepancies. A key issue is the underrepresentation of the CO_2 fertilization effect, a major factor contributing to the increased terrestrial carbon sink over recent decades. This omission could potentially bias our understanding of ecosystem responses to climate change.

Here, we introduce CEDAR-GPP, the first global machine-learning-upscaled GPP product that incorporates the direct CO_2 fertilization effect on photosynthesis. Our product is comprised of monthly GPP estimates and their uncertainty at 0.05º resolution from 1982 to 2020, generated using a comprehensive set of eddy covariance measurements, multi-source satellite observations, climate variables, and machine learning models. Importantly, we used both theoretical and data-driven approaches to incorporate the direct CO_2 effects. Our machine learning models effectively predict monthly GPP ($R^2 \sim 0.72$), the mean seasonal cycles ($R^2 \sim 0.77$), and spatial variabilities ($R^2 \sim 0.63$) based on cross-validation at flux sites. After incorporating the direct CO_2 effects, predicted long-term

GPP trend across global flux towers substantially increases from 3.1 gC m^{-2} year^{-1} to 4.5 – 5.4 gC m^{-2} year^{-1}, which aligns more closely with the 7.7 gC m^{-2} year^{-1} trend detected from eddy covariance data. While the global patterns of annual mean GPP, seasonality, and interannual variability generally align with existing satellite-based products, CEDAR-GPP demonstrates higher long-term trends globally after incorporating CO_2 fertilization and reflected a strong temperature control on direct CO_2 effects. The estimated global GPP trend is 0.57 – 0.76 PgC year^{-1} from 2001 to 2018 and 0.32 – 0.34 PgC year^{-1} from 1982 to ~~2020~~2018. Estimating and validating GPP trends in data-scarce regions, such as the tropics, remains challenging, underscoring the importance of ongoing ground-based monitoring and advancements in modeling techniques. CEDAR-GPP offers a comprehensive representation of GPP temporal and spatial dynamics, providing valuable insights into ecosystem-climate interactions. The CEDAR-GPP product is available at https://zenodo.org/doi/10.5281/zenodo.8212706 (Kang et al., 2024).

1. Introduction

Terrestrial ecosystem photosynthesis, known as Gross Primary Productivity (GPP), is the primary source of food and energy for the Earth system and human society (Keenan and Williams, 2018). Through photosynthesis, terrestrial ecosystems also mitigate climate change, by removing thirty percent of anthropogenic carbon emissions from the atmosphere each year (Friedlingstein et al., 2023). However, due to the lack of direct measurements at the global scale, our understanding of photosynthesis and its spatiotemporal dynamics is limited, leading to considerable disagreements among various GPP estimates (Anav et al., 2015; O'Sullivan et al., 2020; Smith et al., 2016; Yang et al., 2022). Addressing these uncertainties is crucial for improving the predictability of ecosystem dynamics under climate change (Friedlingstein et al., 2014).

Over the past three decades, global networks of eddy covariance flux towers collected *in situ* carbon flux measurements that allow for accurate estimates of GPP, providing valuable insights into photosynthesis dynamics under various environmental conditions (Baldocchi, 2020; Beer et al., 2010). To quantify and understand GPP at scales and locations beyond the ~ $1km^2$ flux tower footprints, machine learning has been employed with gridded satellite and climate datasets to upscale site-based measurements and produce wall-to-wall GPP maps (Dannenberg et al., 2023; Joiner and Yoshida, 2020; Jung et al., 2011; Tramontana et al., 2016; Xiao et al., 2008; Yang et al., 2007; Zeng et al., 2020). This "upscaling" approach provides data-driven and observation-based quantifications without prescribed functional relations between GPP and its climatic or environmental drivers. It offers unique empirical constraints of ecosystem carbon dynamics, complementing those derived from process-based and semi-process-based approaches such as terrestrial biosphere models or the Light Use Efficiency (LUE) models (Beer et al., 2010; Gampe et al., 2021; Jung et al., 2017; Schwalm et al., 2017). In recent years, the growth of global and regional flux networks, coupled with increasing efforts in data standardization, has offered new opportunities for the advancement of upscaling frameworks, enabling comprehensive quantifications of terrestrial photosynthesis (Joiner and Yoshida, 2020; Nelson et al., 2024; Pastorello et al., 2020).

Effective machine learning upscaling depends on a complete set of input predictors that fully explain GPP dynamics. Upscaled datasets have primarily relied on satellite-observed greenness indicators, such as vegetation indices, Leaf Area Index (LAI), the fraction of absorbed photosynthetically active radiation (fAPAR), which effectively capture canopy-level GPP dynamics related to leaf area changes (Joiner and Yoshida, 2020; Ryu et al., 2019; Tramontana et al., 2016).

However, important aspects of leaf-level physiology, such as those controlled by climate factors, are often omitted in major upscaled datasets, preventing accurate characterization of GPP responses to climate change (Bloomfield et al., 2023; Stocker et al., 2019). In particular, none of the previous upscaled datasets have considered the direct effect of atmospheric CO_2 on leaf-level photosynthesis, which is a key factor contributing to at least half of the enhanced land carbon sink observed over the past decades (Keenan et al., 2016, 2023; Ruehr et al., 2023; Walker et al., 2021). This omission can lead to incorrect inferences regarding long-term trends in various components of the terrestrial carbon cycle (De Kauwe et al., 2016).

Multiple independent lines of evidence from the atmospheric inversion (Wenzel et al., 2016), atmospheric $^{13}C/^{12}C$ measurements (Keeling et al., 2017), ice core records of carbonyl sulfide (Campbell et al., 2017), glucose isotopomers (Ehlers et al., 2015), as well as free-air CO_2 enrichment experiments (FACE) (Walker et al., 2021), suggest a widespread positive effect of elevated atmospheric CO_2 on GPP from site to global scales. Increasing atmospheric CO_2 *directly* stimulates the biochemical rate or the light use efficiency (LUE) of leaf-level photosynthesis, known as the direct CO_2 fertilization effect (CFE). Enhanced photosynthesis could lead to greater net carbon assimilation, contributing to an increase in total leaf area. This expansion, contributing to a higher light interception, further enhances canopy-level photosynthesis (i.e. GPP), which is referred to as the indirect CFE. The direct CFE has been found to dominate GPP responses to CO_2 compared to the indirect effect, from both theoretical and observational analyses (Chen et al., 2022; Haverd et al., 2020; Keenan et al., 2023).

Satellite-based estimates have shown an increasing global GPP trend in the past few decades largely attributable to CO_2-induced increases in LAI (Chen et al., 2019; De Kauwe et al., 2016; Piao et al., 2020; Zhu et al., 2016). However, previous upscaled GPP datasets, as well as most LUE models such as the MODIS GPP product, have failed to consider the direct CO_2 effects on leaf-level biochemical processes (Jung et al., 2020; Zheng et al., 2020). Consequently, these products likely underestimated the long-term trend of global GPP, leading to large discrepancies when compared to process-based models, which typically consider both direct and indirect CO_2 effects (Anav et al., 2015; De Kauwe et al., 2016; Keenan et al., 2023; O'Sullivan et al., 2020). Notably, recent improvements in LUE models have included the CO_2 response and show improved long-term changes in GPP globally (Zheng et al., 2020), yet, this important mechanism is still missing in GPP products upscaled from *in situ* eddy covariance flux measurements based on machine learning models.

To improve the quantification of GPP spatial and temporal dynamics and provide a robust representation of long-term dynamics in global photosynthesis, we developed the CEDAR-GPP[1] data product. CEDAR-GPP was upscaled from global eddy covariance carbon flux measurements using machine learning along with a broad range of multi-source satellite observations and climate variables. In addition to incorporating direct CO_2 fertilization effects on photosynthesis, we also account for indirect effects via greenness indicators and include novel satellite datasets such as solar-induced fluorescence (SIF), Land Surface Temperature (LST) and soil moisture to explain variability under environmental stresses. We provide monthly GPP estimations and associated uncertainties at 0.05° resolution derived from ten model setups. These setups differ by the temporal range depending on satellite data availability, the method for incorporating the direct CO_2 fertilization effects, and the partitioning approach used to derive GPP from eddy covariance measurements. Short-term model setups are primarily based on data derived from MODIS satellites generating GPP estimates from 2001 to 2020, while long-term estimates span 1982 to 2020 using combined Advanced Very High Resolution Radiometer (AVHRR) and MODIS data. We used two approaches to incorporate the direct CO_2 fertilization effects, including direct prescription with eco-evolutionary theory and machine learning inference from the eddy-covariance data. Additionally, we provide a baseline configuration that did not incorporate the direct CO_2 effects. Uncertainties in GPP estimation were quantified using bootstrapped model ensembles. We evaluated the machine learning models' skills in predicting monthly GPP, seasonality, interannual variability, and trend against eddy covariance measurements, and compared the CEDAR-GPP spatial and temporal variability to existing satellite-based GPP estimates.

2. Data and Methods

2.1 Eddy covariance data

We obtained monthly eddy covariance GPP measurements from 2001 to 2020 from the FLUXNET2015 (Pastorello et al., 2020), AmeriFlux FLUXNET (https://ameriflux.lbl.gov/data/flux-data-products/), and ICOS Warm Winter 2020 (Warm Winter 2020 Team, 2022) datasets. All data were processed with the ONEFLUX pipeline (Pastorello et al., 2020). Following previous upscaling efforts (Tramontana et al., 2016), we selected monthly GPP data

[1] CEDAR stands for upsCaling Ecosystem Dynamics with ARtificial intelligence

with at least 80% of high-quality hourly or half-hourly data for temporal aggregation. High-quality data refers to GPP derived from measured or high-quality gap-filled Net Ecosystem Exchange (NEE) data. We further excluded large negative GPP values, setting a cutoff of -1 gC m^{-2} day^{-1}. We utilized GPP estimates from both the night-time (GPP_REF_NT_VUT) and day-time (GPP_REF_DT_VUT) partitioning approaches. We classified flux tower sites according to the C3 and C4 plant categories reported in metadata and related publications when available and used a C4 plant percentage map (Still et al., 2003) otherwise. This classification information is included in Supplementary Text S1. Our analysis encompassed 233 sites, predominantly located in North America, Western Europe, and Australia (Figure 1). A list of the sites is provided in Appendix A. Despite their uneven geographical distribution, these sites effectively cover a diverse range of climatic conditions and are representative of global biomes (Figure 1c, 1d). In total, our dataset included over 18000 site-months. Note that we did not include eddy covariance data before 2001, since it was limited to only a few sites with only four sites containing data before 1996. This scarcity might introduce biases in the machine learning models, particularly in the relationship between GPP and CO_2, leading to unreliable extrapolations across space and time in the long-term predictions.

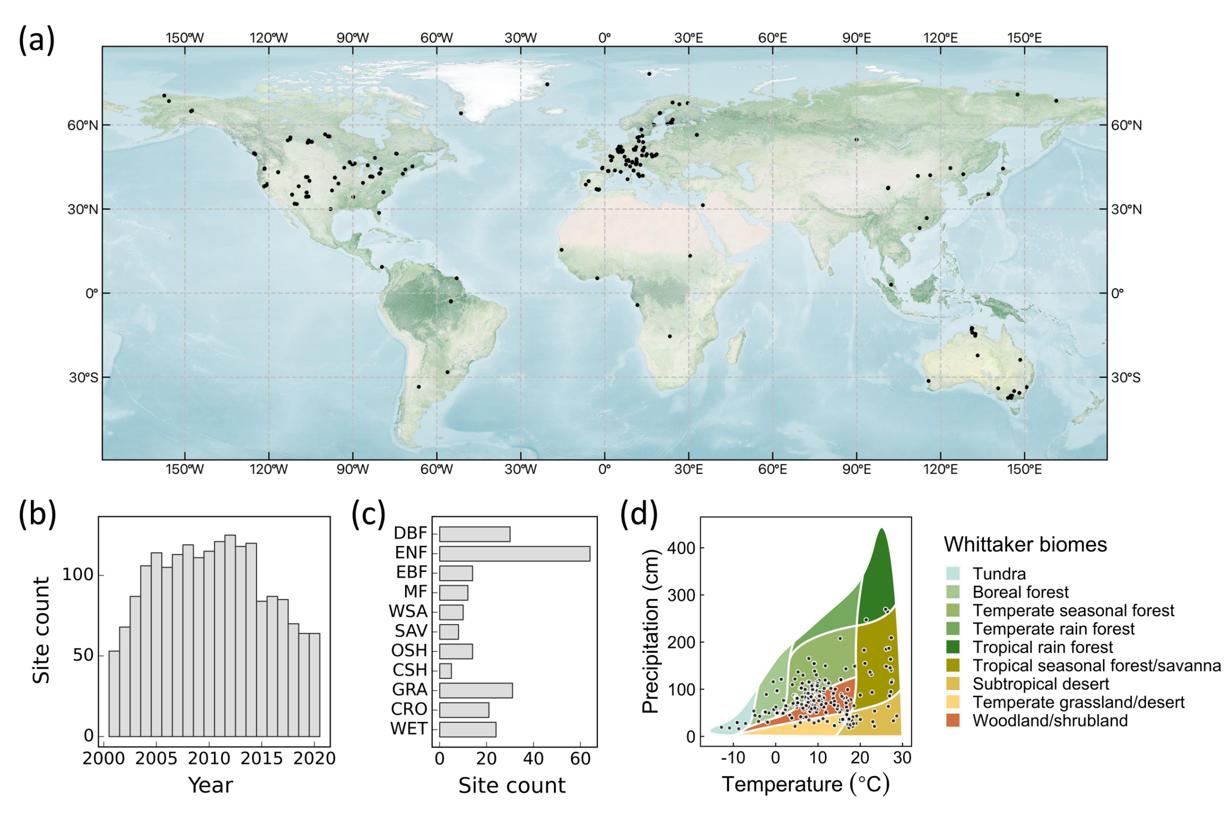


Figure 1. (a) Spatial distribution of eddy covariance sites used to generate the CEDAR-GPP product. (b) Annual site counts. (c) Site counts by biomes. ENF: evergreen needleleaf forests, EBF: evergreen broadleaf forests, DBF: deciduous broadleaf forests, MF: mixed forests, WSA: woody savannas, SAV: savannas, OSH: open shrublands, CSH: closed shrublands, GRA: grasslands, CRO: croplands, WET: wetlands. (d) Sites distributions in the annual temperature and precipitation space. Whittaker biome classification is shown as a reference of natural vegetation based on long-term climatic conditions. It does not directly indicate the actual biome associated with each site. The base map in (a) was obtained from the NASA Earth Observatory map by Joshua Stevens using data from NASA's MODIS Land Cover, the Shuttle Radar Topography Mission (SRTM), the General Bathymetric Chart of the Oceans (GEBCO), and Natural Earth boundaries. Whittaker biomes were plotted using the "plotbiomes" R package (Ștefan and Levin, 2018).

2.2 Global input datasets

We compiled an extensive set of covariates from gridded climate reanalysis data, multi-source satellite datasets including optical, thermal, and microwave observations, as well as categorical information on land cover, climate zone, and C3/C4 classification. The datasets that we compiled offer comprehensive information about GPP dynamics and its responses to climatic variabilities and stresses. Table 1 lists the datasets and associated variables used to generate CEDAR-GPP.

Table 1. Datasets used in different model setups to generate the CEDAR GPP product. Refer to Table S1 for a list of specific variables from each dataset.

Category	Dataset	Temporal coverage	Spatial resolution	Temporal resolution	Usage in model setups		Reference
					Short-term	Long-term	
Climate	ERA5-Land	1950 – present	0.1°	Monthly	✓	✓	(Sabater, 2019)
	ESRA Global Monitoring Laboratory Atmospheric Carbon Dioxide	1976 – present	-	Monthly	✓ (only in CFE-ML and CFE-Hybird setups)	✓ (only in CFE-Hybrid setup)	(Thoning et al., 2021)
Satellite-based datasets	MODIS Nadir BRDF-adjusted reflectance (MCD43C4v006)	2000 – present	0.05°	Daily	✓		(Schaaf and Wang, 2015)
	MODIS Terra and Aqua LAI/fPAR (MCD15A3H, MOD15A2H, v006)	2000 – present	500m	4-day, 8-day	✓		(Myneni et al., 2015a, b)
	MODIS Terra and Aqua LST (MYD11A1, MOD11A1, v006)	2000 – present	1 km	Daily	✓		(Wan et al., 2015b, a)
	BESS_Rad	2000 – 2020	0.05°	Daily	✓		(Ryu et al., 2018)
	Continuous-SIF (from OCO-2 and MODIS)	2000 – 2020	0.05°	4-day	✓		(Zhang, 2021)
	ESA CCI Soil Moisture Combined Passive and Active v06.1	1979 – 2021	0.25°	Daily	✓		(Gruber et al., 2019)
	GIMMS LAI4g	1982 – 2021	0.0833°	Half-month		✓	(Cao et al., 2023)
	GIMMS NDVI4g	1982 – 2021	0.0833 °	Half-month		✓	(Li et al., 2023b)
Static categorical datasets	MODIS Land Cover (MCD12Q1v006)	Average status used between 2001 and 2020	500m	-	✓	✓	(Friedl and Sulla-Menashe, 2019)
	Koppen-Geiger Climate Classification	present	1 km	-	✓	✓	(Beck et al., 2018)
	C4 percentage map	present	1°	-	✓	✓	(Still et al., 2003, 2009)

2.2.1 Climate variables

We obtained air temperature, vapor pressure deficit, precipitation, potential evapotranspiration, and skin temperature from the EAR5-Land reanalysis dataset (Sabater, 2019) (Table 1; Table S1). We applied a three-month lag to precipitation, to ~~reflect the memory of soil moisture and~~ represent the root zone water availability. Averaged monthly atmospheric CO_2 concentrations were calculated as an average of records from the Mauna Loa Observatory and South Pole Observation stations, retrieved from NOAA's Earth System Research Laboratory (Thoning et al., 2021).

2.2.2 Satellite datasets

We assembled a broad collection of satellite-based observations of vegetation greenness and structure, LST, solar radiation, solar-induced fluorescence (SIF), and soil moisture (Table 1, Table S1).

We used three MODIS version 6 products: surface reflectance, LAI/fAPAR, and LST. Surface reflectance from optical to infrared bands (band 1 to 7) was sourced from the MODIS Nadir BRDF-adjusted reflectance (NBAR) daily dataset (MCD43C4) (Schaaf and Wang, 2015). From these data, we derived vegetation indices, including NIRv (Badgley et al., 2019), kNDVI (Camps-Valls et al., 2021), NDVI, Enhanced Vegetation Index (EVI), Normalized Difference Water Index (NDWI) (Gao, 1996), and the green chlorophyll index (CIgreen) (Gitelson, 2003). We also used snow percentages from the NBAR dataset. We used the 4-day LAI and fPAR composite derived from Terra and Aqua satellites (MCD15A3H) (Myneni et al., 2015a; Yan et al., 2016a, b) from July 2002 onwards and the MODIS 8-day LAI and fPAR dataset from Terra only (MOD15A2H) prior to July 2002 (Myneni et al., 2015b). We used day-time and night-time LST from the Aqua satellite (MYD11A1) (Wan et al., 2015b), with the Terra-based LST product (MOD11A1) used after July 2002 (Wan et al., 2015a). Terra LST was bias-corrected with the differences in the mean seasonal cycles between Aqua and Terra following Walther et al. (2022).

We used the PKU GIMMS NDVI4g dataset (Li et al., 2023b) and PKU GIMMS LAI4g (Cao et al., 2023) datasets available from 1982 to 2020. PKU GIMMS NDVI4g is a harmonized time series that includes AVHRR-based NDVI from 1982 to 2003 (with biases and corrections mitigated through inter-calibration with Landsat surface reflectance images) and MODIS NDVI from 2004 onward. PKU GIMMS LAI4g consisted of consolidated AVHRR-based LAI from 1982 to 2003 (generated using machine learning models trained with Landsat-based LAI data and NDVI4g) and reprocessed MODIS LAI (Yuan et al., 2011) from 2004 onwards.

We utilized photosynthetically active radiation (PAR), diffusive PAR, and shortwave downwelling radiation from the BESS_Rad dataset (Ryu et al., 2018). We obtained the continuous-SIF (CSIF) dataset (Zhang, 2021; Zhang et al., 2018) produced by a machine learning algorithm trained using OCO-2 SIF observations and MODIS surface reflectance. We used surface soil moisture from the ESA CCI soil moisture combined passive and active product (version 6.1) (Dorigo et al., 2017; Gruber et al., 2019).

2.2.3 Other categorical datasets

We used plant functional type (PFT) information derived from the MODIS Land Cover product (MCD12Q1) (Friedl and Sulla-Menashe, 2019). We followed the International Geosphere-Biosphere Program classification scheme but merged several similar categories to maximize the amount of eddy covariance sites/observations available for each category. Closed shrublands and open shrublands are combined into a shrubland category. Woody savannas and savannas are combined into savannas. We generated a static PFT map by taking the mode of the MODIS land cover time series between 2001 – 2020 at each pixel to mitigate uncertainties from misclassification in the MODIS dataset. Nevertheless, changes in vegetation structure induced by land use and land cover change are reflected in the dynamics surface reflectance and LAI/fAPAR datasets we used. We used the Koppen-Geiger main climate groups (tropical, arid, temperate, cold, and polar) (Beck et al., 2018). We also utilized a C4 plant percentage map to account for different photosynthetic pathways when incorporating CO_2 fertilization (Still et al., 2003, 2009). The C4 percentage dataset was constant over time.

2.2.4 Data preprocessing

We implemented a three-step preprocessing strategy for the satellite datasets: 1) quality control, 2) gap-filling, and 3) spatial and temporal aggregation. Firstly, we selected high-quality data based on the quality control flags of the satellite products when available. For the MODIS NBAR dataset (MCD43C3), we used data with 75% or more high-resolution NBAR pixels retrieved with full inversions for each band. For MODIS LST, we selected the best quality data from the quality control bitmask as well as data where retrieved values had an average emissivity error of no more than 0.02. For MODIS LAI/fAPAR, we used retrievals from the main algorithm with or without saturation. We used all available data in ESA-CCI soil moisture due to the presence of substantial data gaps. In the gap-filling step, missing values in satellite datasets were temporally filled at the native temporal resolution, following a two-step protocol adapted from Walther et al (2021). Short temporal gaps were first filled with medians from a moving window, and the remaining gaps were filled with the mean seasonal cycle. For datasets with a high temporal resolution, including MODIS NBAR (daily), LAI/fPAR (4-day), BESS (4-day), CSIF (4-day), ESA-CCI (daily), temporal gaps no longer than 5 days (8 days for 4-day resolution products) were filled with medians of 15-day moving windows in the first step. An exception is MODIS LST (daily), for which we used a shorter moving window of 9 days due to rapid changes in surface temperature. GIMMS LAI4g and NDVI4g data were only filled with mean seasonal cycle due to their low temporal resolution (half-month). This is because vegetation structure could experience significant changes at half-month intervals, and gap-filling using temporal medians within moving windows could introduce considerable uncertainties and potentially over-smooth the time series.

Finally, all the datasets were aggregated to a monthly time step and 0.05-degree spatial resolution. We employed the conservative resampling approach using the xESMF python package (Zhuang et al., 2023). To generate the machine learning model training data, we extracted values from the nearest 0.05 degree pixel relative to the site locations within the gridded dataset.

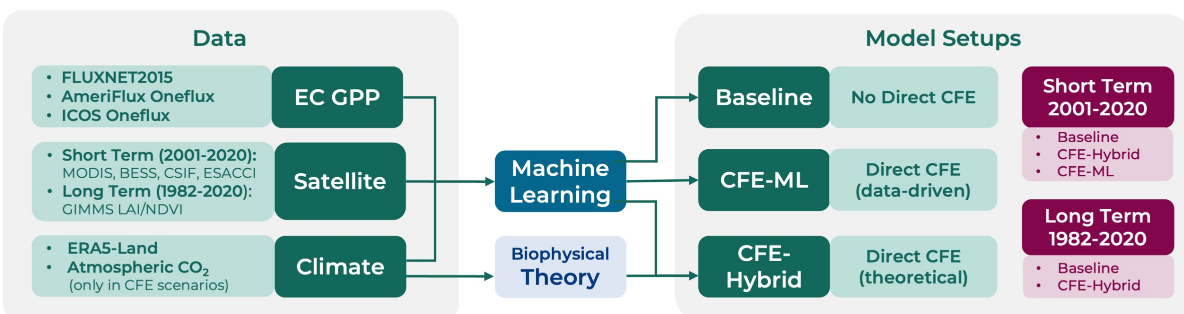


Figure 2. Schematic overview of the CEDAR-GPP model setups.

2.3 Machine learning upscaling

2.3.1 CEDAR-GPP model setups

We trained machine learning models with eddy covariance GPP measurements as targets and climate/satellite variables as input features. We created ten model setups to produce different global monthly GPP estimates (Figure 2; Table 2). The model setups were characterized by the temporal range depending on input data availability, the configuration of CO_2 fertilization effects, and the partitioning approach used to derive the GPP from eddy covariance measurements.

The short-term (ST) model configuration produced GPP from 2001 to 2020, and the long-term (LT) configuration spanned 1982 to 2020. Each temporal configuration uses a different set of input variables depending on their availability. Inputs for the short-term configuration included MODIS, CSIF, BESS PAR, ESA-CCI soil moisture, ERA5-Land, as well as PFT and Koppen Climate zone as categorical variables with one-hot encoding. The long-term used GIMMS NDVI4g and LAI4g data, ERA5-land, PFT and Koppen climate. ESA CCI soil moisture datasets were excluded from the long-term model setups due to concerns about the product quality in the early years when the number and quality of microwave satellite data were limited (Dorigo et al., 2015). A detailed list of input features for each setup is provided in Table S1.

Regarding the direct CO_2 fertilization effects (CFE), we established a "Baseline" configuration that did not incorporate these effects, a "CFE-Hybrid" configuration that incorporated the effects via eco-evolutionary theory, and a "CFE-ML" configuration that inferred the direct effects from eddy covariance data using machine learning. Detailed information about these approaches is provided in Sec. 2.3.2. Furthermore, separate models were trained for GPP target variables from the night-time (NT) and daytime (DT) partitioning approaches.

Table 2 lists the characteristics of ten model setups. Due to the limited availability of eddy covariance observations before 2001, we did not apply the CFE-ML approach to the long-term setups. The CFE-ML model, when trained on data from 2001 to 2020 with atmospheric CO_2 ranging from 370 to 412 ppm, would not accurately predict GPP response to CO_2 for the period 1982 – 2000 when the CO2 levels were markedly lower (340 – 369 ppm). This is because machine learning models, especially tree-based models, could not extrapolate beyond the range of the training data.

Table 2. Specifications of the CEDAR-GPP model setups.

Model Setup Name	Temporal range	Direct CO_2 Fertilization Effects		GPP Partitioning Method
		Configuration	Method	
ST_Baseline_NT	Short-term (ST) 2001 – 2020	Baseline	Not incorporated	Night-time (NT)
ST_Baseline_DT				Day-time (DT)
ST_CFE-Hybrid_NT		CFE-Hybrid	Theoretical	NT
ST_CFE-Hybrid_DT				DT
ST_CFE-ML_NT		CFE-ML	Data-driven	NT
ST_CFE-ML_DT				DT
LT_Baseline_NT	Long-term (LT) 1982 – 2020	Baseline	Not incorporated	NT
LT_Baseline_DT				DT
LT_CFE-Hybrid_NT		CFE-Hybrid	Theoretical	NT
LT_CFE-Hybrid_DT				DT

2.3.2 CO_2 fertilization effect

We established three configurations regarding the direct CO_2 fertilization effects on photosynthesis. In the baseline configuration, we trained machine learning models with eddy covariance GPP~~ measurements~~, input climate and satellite features, but excluding CO_2 concentration. As such, the models only include indirect CO_2 effects from the satellite-based proxies of vegetation greenness or structure representing changes in canopy light interception, and they do not consider the direct effect of CO_2 on leaf-level photosynthetic rates (or light use efficiency, LUE). Our baseline model is therefore directly comparable to other satellite-derived GPP products that only account for indirect CO2 effects (Joiner and Yoshida, 2020; Jung et al., 2020).

In the CFE-ML configuration, we added monthly CO_2 concentration into the feature set in addition to those incorporated in the baseline models. Models inferred the functional relationship between GPP and CO_2 from the eddy covariance data. They thus encompass both CO_2 fertilization pathways – direct effects on LUE and indirect effects from the satellite-based proxies of vegetation greenness and structure.

In the CFE-Hybrid configuration, we applied biophysical theory to estimate the response of LUE to elevated CO_2, i.e. the direct CFE (Appendix B~~A~~). First, we estimated a reference GPP, where LUE was not affected by any increase in atmospheric CO_2, by applying the CFE-ML model with a constant atmospheric CO_2 concentration equal to the 2001 level while keeping all other variables temporally dynamic. Then, the impacts of CO_2 on LUE were prescribed onto the reference GPP estimates using a theoretical CO_2 sensitivity function of LUE according to the optimal coordination theory (Appendix ~~A~~B). The theoretical CO_2 sensitivity function represents a CO_2 sensitivity that is equivalent to that of the electron-transport-limited (light-limited) photosynthetic rate. When light is limited, elevated CO_2 suppresses photorespiration leading to increased photosynthesis at a lower rate than when photosynthesis is limited by CO_2 (Lloyd and Farquhar, 1996; Smith and Keenan, 2020). Thus, the CFE-Hybrid scenario provides a conservative estimation of the direct CO_2 effects on LUE. Note that the theoretical sensitivity function describes the fractional change in LUE due to direct CO_2 effects relative to a reference period (i.e. 2001). Therefore, we used the CFE-ML model to establish this reference GPP by fixing the CO_2 effects to the 2001 level, rather than simply using the GPP from the Baseline model in which the direct CO_2 effects were not represented. Long-term trends from the reference and the Baseline models are consistent.

For both CFE-ML and CFE-Hybrid scenarios, we made another conservative assumption that C4 plants do not benefit from elevated CO_2, despite potential increases in photosynthesis during water-limited conditions due to enhanced water use efficiency (Walker et al., 2021). Data from flux tower sites dominated by C4 plants were removed from our training set, so the machine learning models inferred CO_2 fertilization only from flux tower sites dominated by C3 plants. When applying models globally, we assumed the reference GPP values (with constant atmospheric CO_2 concentration equal to the 2001 level) to represent C4 plants, and GPP estimates from CFE-ML or CFE-Hybrid models were applied in proportion to the percentage of C3 plants in a grid cell.

2.3.3 Machine learning model training and validation

We employed the state-of-the-art XGBoost machine learning model, known for its high accuracy in regression problems across various domains, including environmental and ecological predictions (Berdugo et al., 2022; Chen and Guestrin, 2016; Kang et al., 2020). XGBoost is a scalable and parallelized implementation of the gradient boosting technique that iteratively trains an ensemble of decision trees, with each iteration targeting to minimize the residuals from the last iteration. A notable merit of XGBoost is its ability to make predictions in the presence of missing values, a common issue in remote sensing datasets. ~~Without relying on prior assumptions about the functional forms or statistical distributions, t~~The model is also robust to multi-collinearity between the predictors in our dataset, particularly for the variables derived from MODIS data.

We used five-fold cross-validation for model evaluation. Training data was randomly split into five groups (folds), with each fold held out for testing while the rest four folds were used for model training. We imposed two restrictions on fold splitting: each flux site was entirely assigned to a fold to test model performance over unseen locations; the random sampling was stratified based on PFT to ensure coverage of the full range of PFTs in both training and testing. Additionally, co-located sites, defined as those within 0.05° of each other, were also assigned to the same fold, as they were often setup as a cluster with different treatments. This approach avoids conflated estimates of model uncertainty, as these sites are not independent. We also used a nested-cross-validation strategy, during which we performed a randomized search of hyperparameters using three-fold cross-validation within the training set. The nested-cross-validation was aimed to reduce the risk of overfitting and improve the robustness of the evaluation.

We assessed the models' ability to capture the temporal and spatial characteristics of GPP, including monthly GPP, mean seasonal cycles, monthly anomalies, and cross-site variability. Model performance was assessed separately for each model setup (Table 2) and summarized by PFT and Koppen climate zone. Mean seasonal cycles were calculated as the mean monthly GPP over the site observation period, and monthly anomalies were the residuals of monthly GPP after subtracting mean seasonal cycles. Monthly GPP averaged over years for each site was used to assess cross-site variability. Goodness-of-fit metrics include RMSE, bias, and coefficient of determination (R^2).

To evaluate the models' ability to capture long-term GPP trends, we aggregated the monthly GPP to annual values following Chen et al. (2022) , which detected the CO_2 fertilization effect across global eddy covariance sites. For sites with at least five years of observations, GPP anomalies were computed by subtracting the multi-year mean GPP from the annual GPP for each site. Anomalies were aggregated across sites to achieve a single multi-site GPP anomaly per year. We excluded a site-year if less than 11 months of data was available and used linear interpolation to fill the remaining temporal gaps. This resulted in 81 sites used in the GPP trend evaluation. We used the Sen slope and Mann-Kendall test to examine the GPP trends from 2002 to 2019, excluding 2001 and 2020, due to the limited number of available sites with more than five years of data. We further assessed the aggregated annual trend by grouping the sites based on plant functional types and the Koppen climate zones. Categories with less than six long-term sites available were excluded from the analysis, which includes EBF and Tropics.

To further analyze GPP responses to CO_2 in the CFE-ML models, we leveraged two explainable machine learning approaches: ALE (Accumulated Local Effects) (Apley and Zhu, 2020; Baniecki et al., 2021) and SHAP (SHapley Additive exPlanations) (Lundberg and Lee, 2017). SHAP is a model interpretation method derived from game theory, providing a value for each feature's contribution to a prediction, elucidating how each feature impacts the model's output in a specific instance. Conversely, ALE quantifies the average effect of a feature across the data, isolating its impact by aggregating local effects and avoiding the biases associated with correlated features.

2.3.4 Product generation and uncertainty quantification

In the CEDAR-GPP product, we generated GPP estimates from ~~each of the~~ ten model setups, by applying the model to global gridded datasets ~~within the corresponding temporal range~~ (Table 2). GPP estimates were named after the corresponding model setups. We used bootstrapping to estimate prediction uncertainties. For each model setup, we generated 30 bootstrapped sample sets of eddy covariance data, which were then used to train an ensemble of 30 XGBoost models. The bootstrapping was performed at the site level, and each bootstrapped sample set contained around 140 to 150 unique sites, 17000 to 19000 site months covering all PFTs. The relative PFT composition in the bootstrapped sample sites was consistent with the full dataset. Hyperparameters of the XGBoost models used in the final product generation were described in Supplementary Text S2. The 30 models trained with bootstrapped samples generated an ensemble of 30 GPP values. We provided the ensemble GPP mean and used standard deviation to indicate uncertainties, for each of the ten model setups.

2.4 Product inter-comparison

We compared the global spatial and temporal patterns of CEDAR-GPP with other major satellite-based GPP products, including three machine learning upscaled and two LUE-based datasets. We obtained two FLUXCOM products (Jung et al., 2020), the latest version of FLUXCOM-RS (FLUXCOM-RSv006) available from 2001 to 2020 based on remote sensing (MODIS collection 6) datasets only, as well as the FLUXCOM-RS+METEO ensemble available between 1979 to 2018 and based on the climatology of remote sensing observations and ERA5 forcings (hereafter FLUXCOM-ERA5). We used FluxSat (Joiner and Yoshida, 2020), available from 2001 to 2019, which is an upscaled dataset based on MODIS NBAR surface reflectance and PAR from Modern-Era Retrospective analysis for Research and Applications 2 (MERRA-2). Importantly, FluxSat does not incorporate climate forcings. We used the MODIS GPP product (MOD17) available since 2001, which was generated based on MODIS fAPAR and LUE as a function of air temperature and vapor pressure deficit but not atmospheric CO_2 concentration (Running et al., 2015). We also used the rEC-LUE products, available from 1982 to 2018 and based on a revised LUE model that incorporated the effect of atmospheric CO_2 concentration and the fraction of diffuse PAR on LUE (Zheng et al., 2020). Additionally, to evaluate GPP trends, we ~~compared our product against~~further included three process-based models forced by remote sensing data – BEPS (Leng et al., 2024), BESSv2 (Li et al., 2023a), and PML V2 (Zhang et al., 2019). These products estimate GPP by scaling leaf-level biochemical photosynthesis models to the canopy level, using satellite-derived vegetation structural variables such as LAI. All three products incorporate the direct CO_2 effects within their biochemical photosynthesis models.

All datasets were resampled to 0.1 ° spatial resolution, and a common mask for the vegetated land area was applied. We evaluated global mean annual GPP, mean seasonal cycle, interannual variability, and trend among different datasets, comparing them over a common time period determined by their data availability. Global total GPP was computed by scaling the global area-weighted average GPP flux with the global land area (122.4 million km^2) following Jung et al. (2020). Mean seasonal cycle was defined as above (Sec. 2.3.3). We used the standard deviation of annual GPP to indicate the magnitude of interannual variability, the Sen slope to indicate the GPP annual trend, and the Mann-Kendall test for the statistical significance of trends.

3. Results

3.1 Evaluation of model performance

3.1.1 Overall performance

The short-term and long-term models explain approximately 72% and 67%, respectively, of the variation in monthly GPP across global eddy covariance sites (Figure 3a). The long-term models consistently yield lower performance than the short-term models, likely due to differences in the satellite remote sensing datasets used, as the short-term models benefited from richer information from surface reflectance of individual bands, LST, CSIF, as well as soil moisture, while the long-term model only exploited NDVI and LAI. The models with different CFE configurations and target GPP variables (i.e. partitioning approaches) have similar performance in predicting monthly GPP (Figure 3b, Table S2). All models exhibit minimal bias of less than 0.1.

Model performance in terms of the different temporal and spatial characteristics of monthly GPP is variable (Figure 3c-h). The models are most successful at predicting mean seasonal cycles, with the short-term and long-term models explaining around 77% and 72% of the variability, respectively (Figure 3c-d). The short-term and long-term models capture 63% and 54% , respectively, of the spatial variabilities in multi-year mean GPP across global sites (i.e., cross-site variability) (Figure 3g-h). However, all models underestimate monthly anomalies across the sites, with R^2 values below 0.12 (Figure 3e-f). Patterns from the DT setups do not significantly differ from those of the NT setups (Figure S1, Table S2). Model performance also varies across sites, and models are more advantageous in explaining mean seasonal cycles than monthly anomalies in most sites (Figure S2).

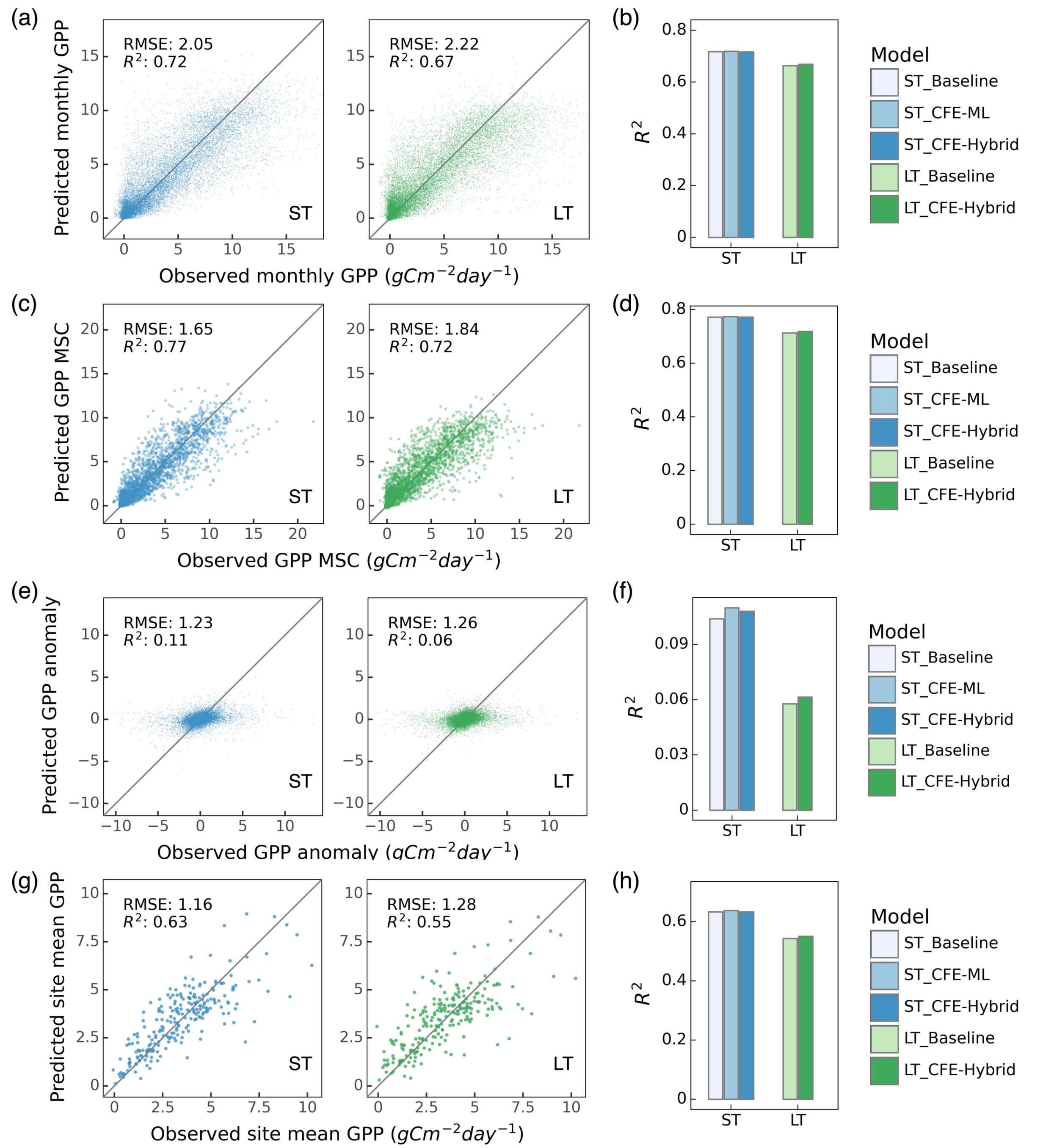


Figure 3. Machine learning model performance in predicting monthly GPP and its spatial and temporal variability. Only NT models are shown and DT results are provided in Supplementary Figure S1. Scatter plots illustrated relationships between model predictions and observations for monthly GPP (a), mean seasonal cycles (MSC) (c), monthly anomaly (e), and cross-site variability (g) for ST_CFE-Hybrid_NT (left, blue) and LT_CFE-Hybrid_NT (right, green) models. Corresponding bar plots show the R^2 values for five NT model setups in predicting monthly GPP (b), MSC (d), monthly anomaly (f), and cross-site variability (h).

3.1.2 Performance by biome and climate zone

The predictive ability of our models varies across different PFTs and Koppen climate zones (Figure 4). Here we present results from the CFE-Hybrid LT and ST models based on NT partitioning and note that patterns for the other CFE configurations and the DT GPP were similar (Figure S3, S4, S5).

Model performance in terms of monthly GPP is the highest for deciduous broadleaf forests, mixed forests, and evergreen needleleaf forests, with R^2 values above 0.76. Model accuracies are also high for savannas and grasslands, followed by croplands and wetlands, with R^2 values between 0.48 and 0.76. Model accuracies are lowest in evergreen broadleaf forests and shrublands, with R^2 values as low as 0.13. Across climate zones, models achieve the highest accuracy in predicting monthly GPP in cold climates with R^2 around 0.73 – 0.78, followed by tropics and temperate zones (R^2 ~ 0.47 – 0.65). The short-term models have the lowest performance in polar regions with an R^2 value of around 0.37, and the long-term models have the lowest performance in arid regions with an R^2 value of 0.28. Interestingly, short-term and long-term models exhibite substantial differences in arid regions and shrublands marked by strong seasonality and interannual variabilities.

Model performance in terms of mean seasonal cycles across PFTs and climate zones follows patterns for monthly GPP, while disparities emerge for performance in terms of GPP anomaly and cross-site variability (Figure 4, Figure S3, S4, S5). The short-term model shows the highest predictive power in explaining monthly anomalies in arid regions with an R^2 value of 0.48, where savanna and shrublands sites are primarily located. Model performance in all other climate zones is significantly lower. The short-term model also demonstrates good performance in capturing anomalies in deciduous broadleaf forests. The long-term model's relative performance between PFTs and climate zones is mostly consistent with that of the short-term model, with lower accuracy in shrublands when compared to the short-term model.

Models demonstrate the highest accuracy in predicting cross-site variability in savannas, grasslands, evergreen needleleaf forests, and evergreen broadleaf forests ($R^2 > 0.36$) and the lowest accuracy in deciduous broadleaf forests, mixed forests, and croplands ($R^2 < 0.1$). The short-term model additionally shows good performance in shrublands and wetlands ($R^2 > 0.36$), whereas the long-term model fails to capture any variability for shrublands. In terms of climate zones, models are most successful at explaining the variabilities within tropical and cold climate zones ($R^2 > 0.50$), the short-term model has moderate performance in temperature and polar regions (R^2 ~ 0.22), and the long-term model has low performance for both temperate and arid regions with R^2 values below 0.16.

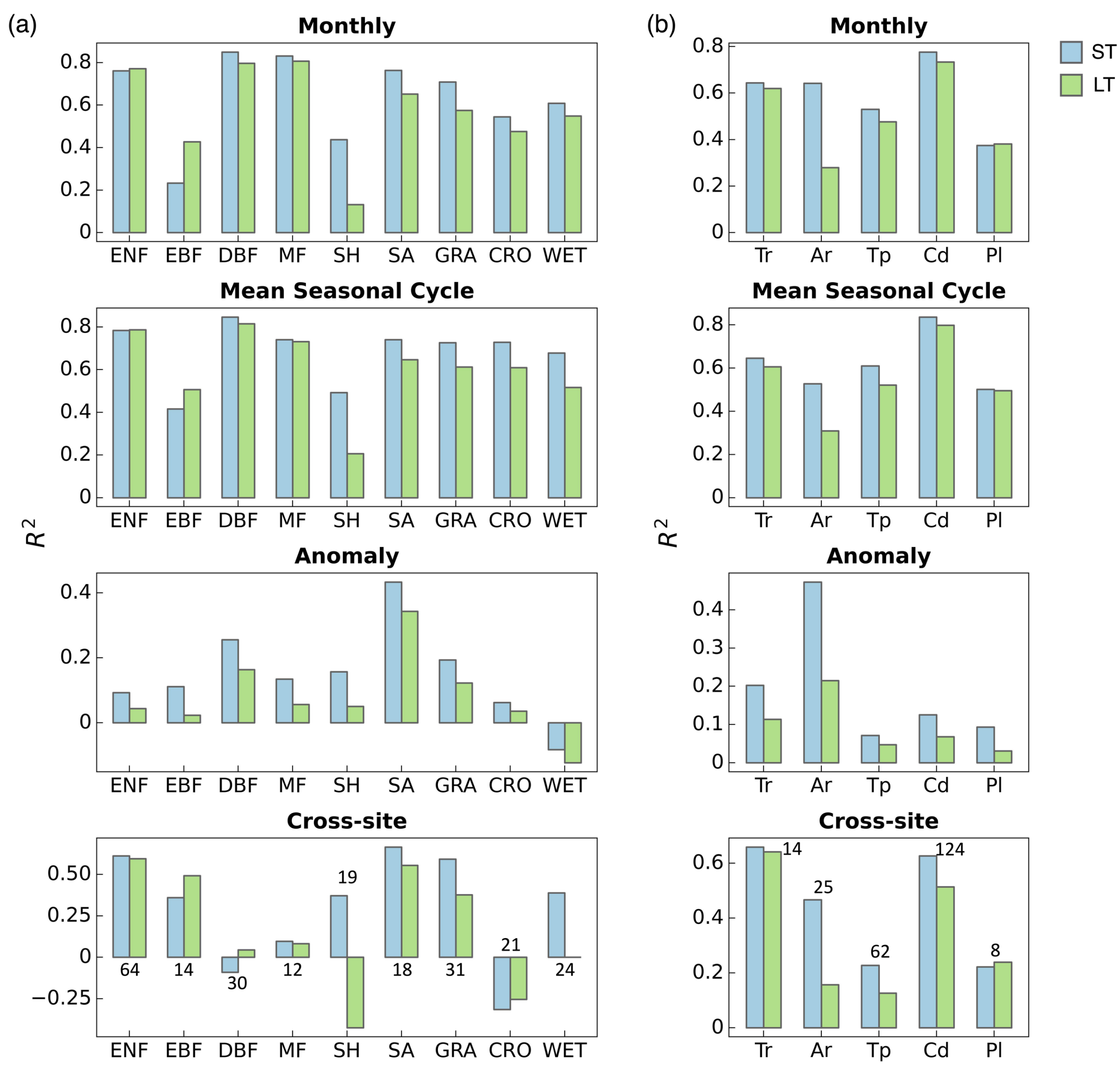


Figure 4. Performance of the ST_CFE-Hybrid_NT (blue) and LT_CFE-Hybrid_NT (green) models on GPP spatiotemporal estimation by plant functional types (a) and climate zones (b). The cross-site panels included the number of sites within each category. Color indicates short-term (ST) or long-term (LT) models. ENF: evergreen needleleaf forest, EBF: evergreen broadleaf forest, DBF: deciduous broadleaf forest, MF: mixed forest, SH: shrubland, SA: savanna, GRA: grassland, CRO: cropland, WET: wetland. Tr: tropical, Ar: arid, Tp: temperate, Cd: cold, Pl: polar. The performance of DT models is displayed in Supplementary Figure S3.

3.1.3 Prediction of long-term trends

Eddy covariance derived GPP presents a substantial increasing trend across flux sites between 2002 and 2019 (Figure 5a, Figure S6a). The eddy covariance GPP from the night-time partitioning approach indicates an overall trend of 7.7 gC m^{-2} $year^{-2}$. In contrast, the ST_ Baseline_NT model predictes a more modest overall trend of 3.1 gC m^{-2} $year^{-2}$ across the flux sites, primarily reflecting the indirect CO_2 effect manifested through the growth of LAI. Both the ST_CFE-ML_NT and ST_CFE-hybrid_NT models predicte much higher trends of 5.4 and 4.5 gC m^{-2} $year^{-2}$ respectively, representing an improvement from the Baseline model by 74% and 45%, aligning more closely to eddy covariance observations. Similarly, the LT_CFE-Hybrid_NT model shows an improved trend estimation than the LT_Baseline_NT model. All trends were statistically significant ($p < 0.05$).

Aggregated eddy covariance GPP experiences increasing trends of varied magnitudes across different climate zones and plant functional types (Figure 5b,c; Figure S6b,c). While the machine learning models generally do not fully capture the enhancement in GPP for most categories, the CFE-ML and/or CFE-hybrid models consistently outperform the Baseline models in both ST and LT setups. The CFE-ML setup predicts a higher trend than CFE-hybrid in most cases, suggesting that the data-driven approach captures more dynamics not represented in the theoretical model, which is based on conservative assumptions regarding the CO_2 sensitivity of photosynthesis (see Sect. 2.3.2 and Appendix ~~A~~B). The choice of remote sensing data (ST vs. LT configurations) does not lead to substantial differences in the predicted GPP trend. Most long-term flux sites (at least 10 years of records) with a significant trend experienced an increase in GPP, and the CFE-ML and/or CFE-hybrid models align closer to eddy covariance data than the Baseline models (Figure S7). Additionally, we found a considerably higher trend in eddy covariance GPP measurements derived from the day-time versus night-time partitioning approach, potentially associated with uncertainties in GPP partitioning methods (Figure S6). Yet, machine learning model predicted trends are not strongly affected by GPP partitioning methods.

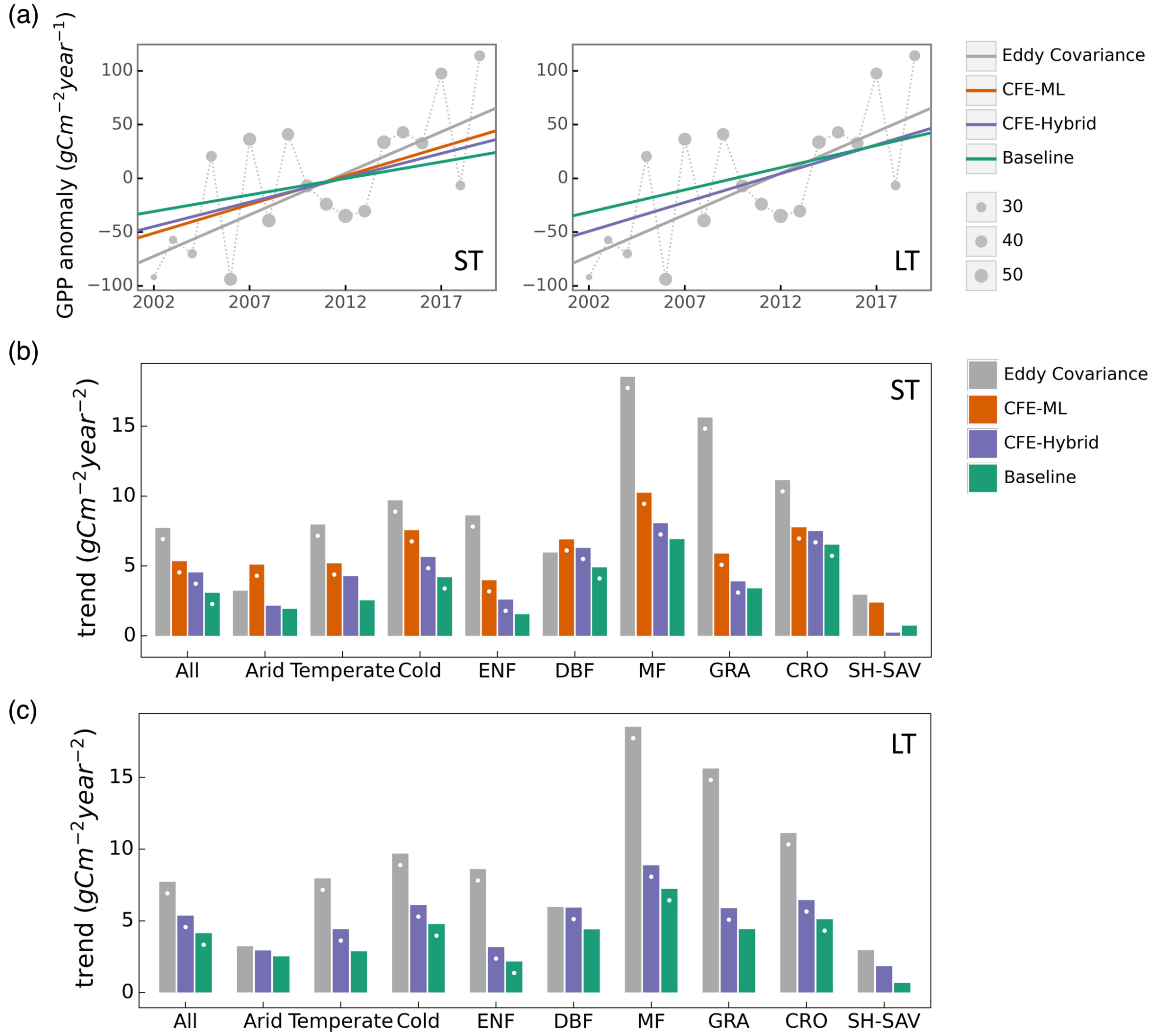


Figure 5. Comparison of observed and predicted GPP (from NT models only) trends across eddy covariance flux towers. (a) Aggregated annual GPP anomaly from 2002 to 2019 and trend lines from eddy covariance (EC) data, and three CFE model setups (short-term, night-time partitioning) for ST (left) and LT (right) models. The size of grey circle markers is proportional to the number of sites. (b) Comparison of annual GPP trends from eddy covariance measurements and the short-term (ST) CEDAR-GPP model setups by plant functional types and climate zones. (c) Comparison of annual GPP trends from eddy covariance measurements and the long-term (LT) CEDAR-GPP model setups by plant functional types and climate zones. In (b) and (c), Categories with less than 6 sites, including Tropics and EBF, were not shown. While dots on the bars indicate statistically significant trends with p-value < 0.1. Results for the DT models are shown in Supplementary Figure S3.

The differences in estimated GPP trends between the Baseline and CFE models underscore the significant long-term GPP changes driven by the direct CO_2 effect. Using explainable machine learning approaches – ALE and SHAP, we further assessed the CFE-ML models for quantifying the direct CO_2 effect. Both approaches reveal~~ed~~ a consistently positive influence of CO_2 on GPP, aligning with biophysical theories (Figure S8). Compared to the effects from light (PAR) and vegetation structures (e.g. NIRv), the impacts of CO_2 are considerably smaller, which explains the minimal differences in overall model accuracy between the Baseline and CFE models.

Finally, we evaluate~~d~~ CEDAR-GPP using independent eddy covariance data (11 sites, Table S3) that was not involved in model training and obtained from the OzFlux FluxNet dataset (Ozflux, 2024). Among these sites, only two - AU-Cpr (Tropical) and AU-Stp (Aird) - with more than five years of records exhibite a GPP trend with p-value less than 0.3. CEDAR-GPP shows strong consistency with the observed trend (Figure S9). Additionally, CEDAR-GPP achieves reasonable accuracy in predicting monthly GPP ($R^2 \sim 0.73 – 0.75$), mean seasonal cycle ($R^2 \sim 0.74 – 0.78$), and monthly anomalies ($R^2 \sim 0.26 – 0.50$) (Table S4, Figure S10), closely aligning with the cross-validation results.

3.2 Evaluation of GPP spatial and temporal dynamics

We compared CEDAR-GPP estimates with other upscaled or LUE-based datasets regarding the mean annual GPP (Sect. 3.2.1), GPP seasonality (Sect. 3.2.2), interannual variability (Sect. 3.2.3), and annual trends (Sect. 3.2.4). CEDAR-GPP model setups generally show similar patterns in mean annual GPP, seasonality, and interannual variability, therefore, in corresponding sections, we present the CFE-Hybrid model setups as representative examples for comparisons with other datasets, unless otherwise stated. Supplementary figures include comparisons involving CEDAR-GPP estimates from all model setups.

3.2.1 Mean annual GPP

Global patterns of mean annual GPP are generally consistent among CEDAR-GPP model setups, FLUXCOM, FLUXSAT, MODIS, and rEC-LUE, with few noticeable regional differences (Figure 6, Figure S11). Differences among CEDAR-GPP model setups are minimal and only evident between the NT and DT setups in the tropics (Figure 6b-c, Figure S11). CEDAR-GPP short-term datasets show highest consistency with FLUXSAT in terms of mean annual GPP magnitudes (2001 – 2018) and latitudinal variations, although FLUXSAT presents slightly higher GPP values in the tropics compared to CEDAR-GPP (Figure 6b). Mean annual GPP magnitudes for FLUXCOM-RS006 and MODIS are lower globally than CEDAR-GPP and FLUXSAT, with the most pronounced differences observed in the tropical areas. Among the long-term datasets (CEDAR-GPP LT, FLUXCOM-ERA5, and rEC-LUE), mean annual GPP (1982 – 2018) exhibites greater disparities in the northern mid-latitudes than in the tropics and southern hemisphere (Figure 6c). CEDAR-GPP aligns more closely with FLUXCOM-ERA5 than with rEC-LUE, with the latter showing lower annual mean GPP globally, particularly between 20ºN to 50º N.

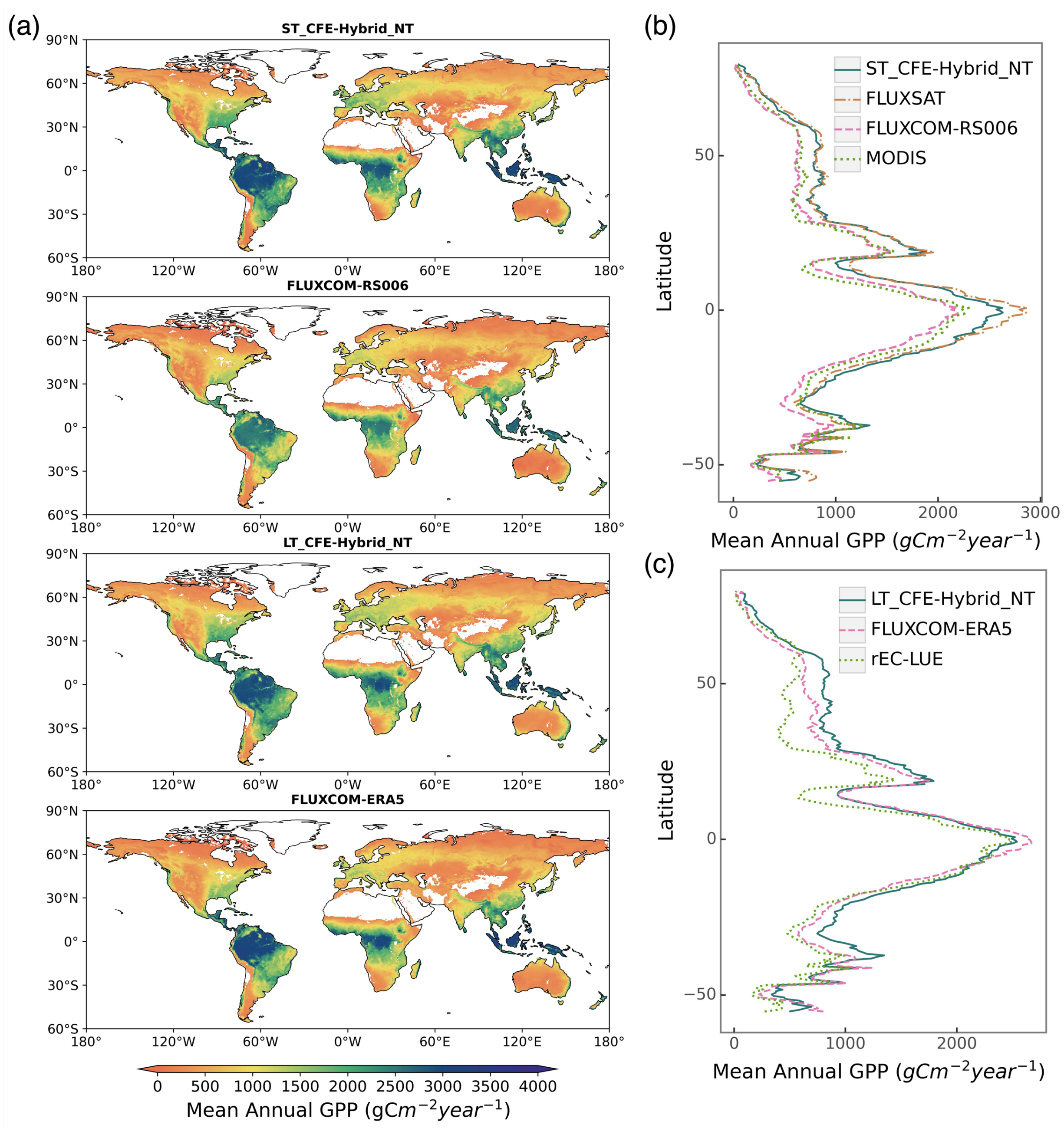


Figure 6. Global distributions of mean annual GPP from CEDAR-GPP and other machine learning upscaled and LUE-based reference datasets. (a) Global patterns of mean annual GPP from two short-term datasets including ST_CFE-Hybrid_NT, and FLUXCOM-RS006, and two long-term datasets including LT_CFE-Hybrid_NT, and FLUXCOM-ERA5. (b) Latitudinal distributions of mean annual GPP from short-term

datasets (ST_CFE-Hybrid_NT, FLUXSAT, FLUXCOM-RS006, and MODIS). (c) Latitudinal distributions of mean annual GPP from long-term datasets (LT_CFE-Hybrid_NT, FLUXCOM-ERA5, and rEC-LUE). Mean annual GPP was computed between 2001 and 2018 for short-term datasets and between 1982 and 2018 for long-term datasets.

3.2.2 Seasonal variability

CEDAR-GPP agrees with other GPP datasets on seasonal variabilities (average between 2001 and 2018) at the global scale, characterized by a peak in GPP in July and a nadir between December and January (Figure 7, Figure S12). At the global scale, CEDAR-GPP is most closely aligned with FLUXSAT in GPP seasonal magnitude and amplitude, while both FLUXCOM and MODIS display a relatively less pronounced magnitude.

In boreal and temperate regions of the Northern Hemisphere, all datasets agree on seasonal GPP variation, with only minor variances in the magnitude of peak GPP. In Southern Hemisphere temperate regions, datasets demonstrate similar seasonality, though with greater variability in peak amplitudes compared to the Northern Hemisphere. The largest disparities are found in the South American tropical areas, where seasonal variation is less prominent. Here, FLUXSAT shows a distinct bi-modal pattern with peaks in March-April and September-October. CEDAR-GPP and FLUXCOM-ERA5 aligns with the second peak, but exhibite a less pronounced first peak. Interestingly, the DT setups of CEDAR-GPP show slightly higher peaks in March-April in this region (Figure S13). MODIS, in contrast, indicates an inverse seasonal pattern with a small peak from June to August. Across all regions, CEDAR-GPP's seasonality aligns more closely with FLUXSAT and FLUXCOM-ERA5 than with other datasets. Differences among the ten CEDAR-GPP model setups are minimal, except for small variations in GPP magnitude in some tropical areas between NT and DT setups (Figure S13).

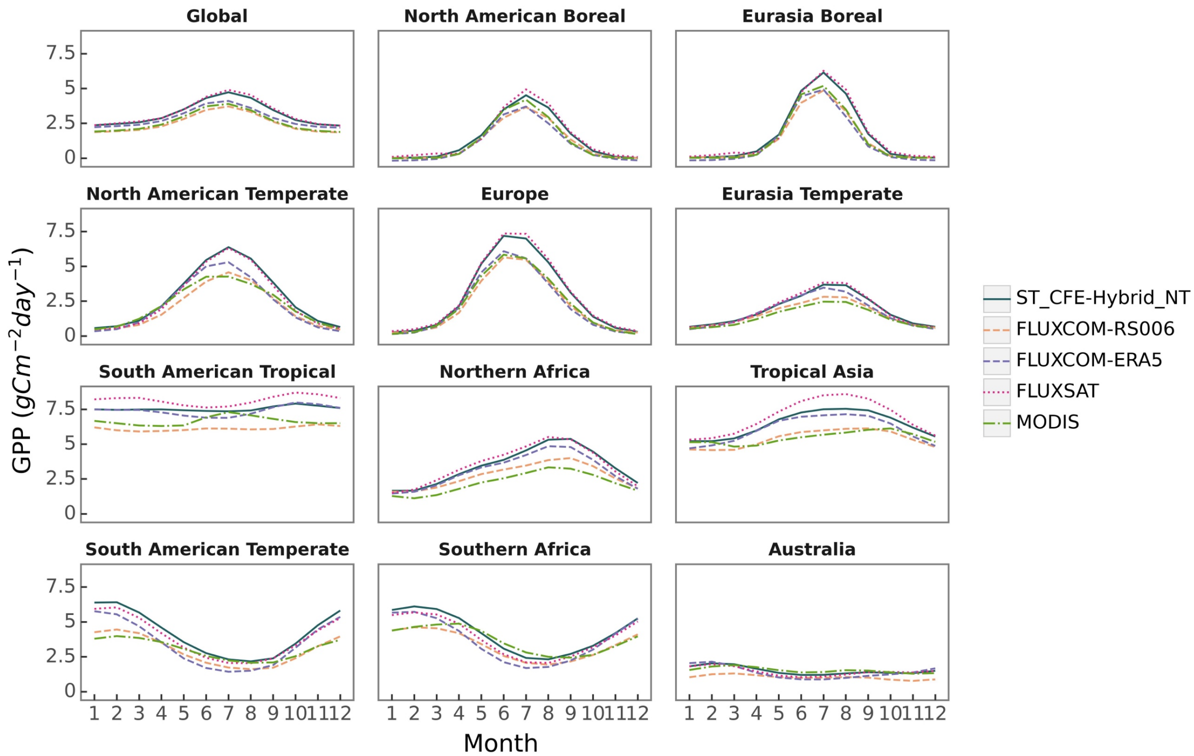


Figure 7. Comparison of global and regional GPP mean seasonal cycle between different datasets on a global scale. Monthly means were averaged from 2001 to 2018 for all datasets. Geographic boundaries of the 11 TransCom land regions were obtained from the CarbonTracker (CT2022) dataset and shown in Figure ~~S19~~S18.

3.2.3 Interannual variability

We found distinct spatial patterns in GPP interannual variability between upscaled and LUE-based datasets and a high level of agreement within each category, with the exception of FLUXCOM-ERA5, which show minimal interannual variability globally (Figure 8, Figure S14). All datasets agree on the presence of GPP interannual variability hotspots in eastern and southern South America, central North America, southern Africa, and western Australia. These hotspots primarily correspond to arid and semi-arid areas characterized by grasslands, shrubs, and croplands (Figure 9). CEDAR-GPP is highly consistent with FLUXSAT, and both datasets also display relatively high interannual variability in the dry subhumid areas of Europe, predominantly covered by croplands. FLUXCOM-RS006 mirrors the relative spatial patterns of CEDAR-GPP and FLUXSAT, albeit at lower magnitudes. The LUE-based datasets (MODIS and rEC-LUE) predict a much higher interannual variability than the upscaled datasets in the tropical areas, particularly in evergreen broadleaf forests and woody savannas (Figure 8, Figure 9). These datasets also depict slightly higher interannual variability for other types of forests, including evergreen needleleaf forests and deciduous broadleaf forests, compared to the upscaled datasets. The lack of interannual variability in FLUXCOM-ERA5 is attributable to the use of mean seasonal cycles of remotely sensed vegetation greenness indicators rather than their dynamic time series. Ten CEDAR-GPP model setups present consistent patterns in interannual variability, and differences were minimal (Figure S14).

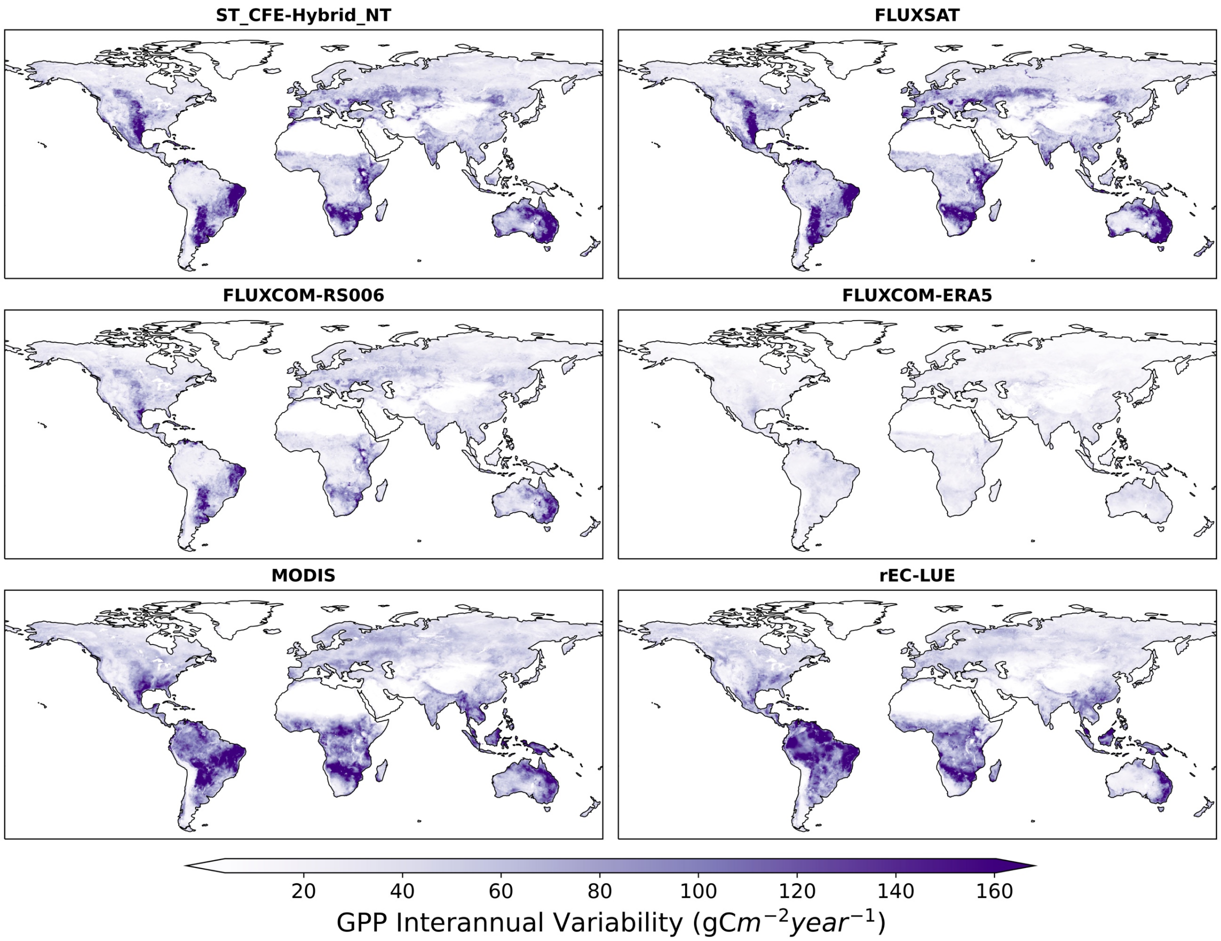


Figure 8. Spatial patterns of GPP interannual variability extracted over 2001 to 2018 for CEDAR-GPP (ST_CFE-Hybrid_NT), FLUXSAT, FLUXCOM-RS006, MODIS, FLUXCOM-ERA5, and rEC-LUE.

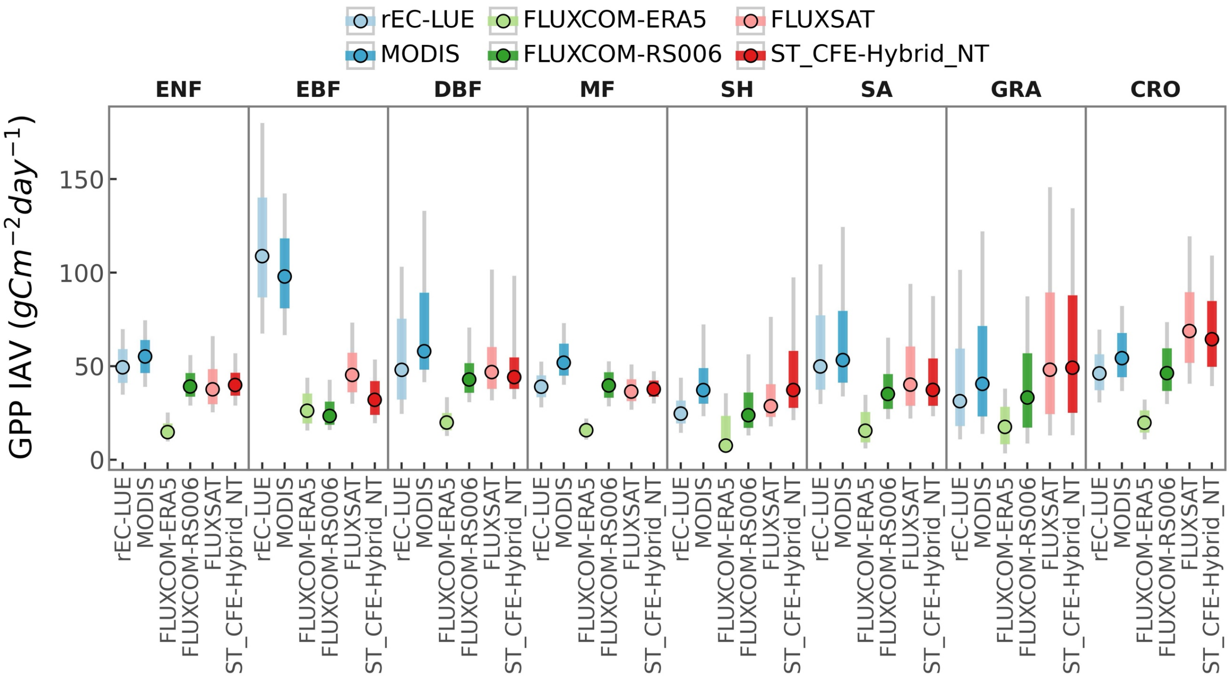


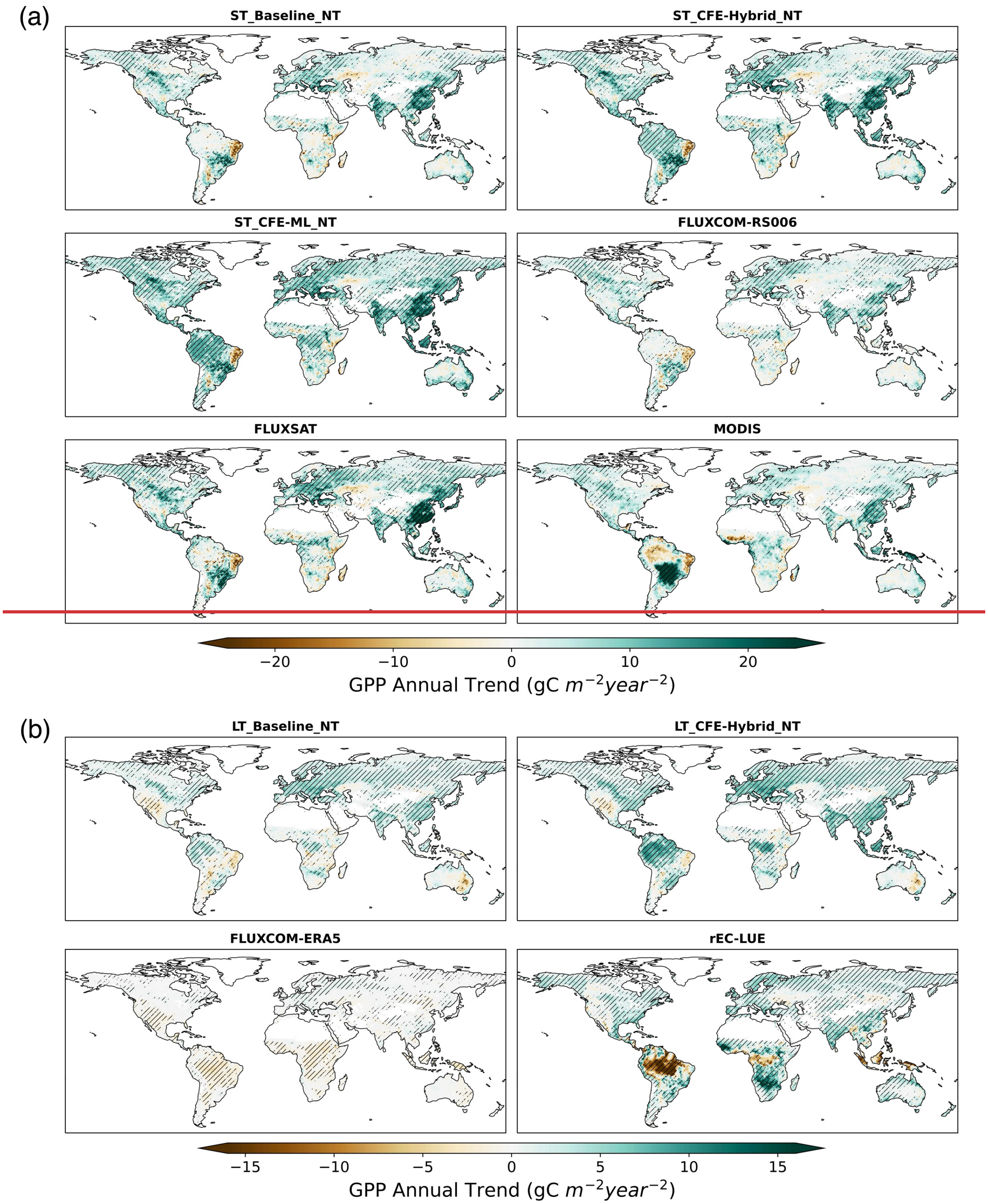
Figure 9. Comparison of GPP interannual variability (IAV) across global datasets by PFT. Colored dots represent the median IAV, thicker gray bars indicate the 25% to 75% percentiles of IAV distributions, and thinner grey bards show the 10% to 90% percentiles.

3.2.4 Trends

Differences in annual GPP trends among CEDAR-GPP model setups and other upscaled and LUE-based datasets mainly reflect the variability in the representation of CO_2 fertilization effects (Figure 10, 11, Figure S15). From 2001 to 2018, the CEDAR-GPP Baseline model setups show spatial variations in GPP trends consistent with the other upscaled datasets without direct CO_2 fertilization effects, including FLUXSAT and FLUXCOM-RSv006. In these datasets, substantial increases are seen in southeastern China and India, western Europe, and part of North and South America. These increases are largely associated with rising LAI due to land use changes and indirect CO_2 fertilization effects, as identified by previous studies (Chen et al., 2019; Zhu et al., 2016). Although MODIS, which also does not include a direct CO_2 fertilization effect, generally agrees with these increasing trends, it ~~also~~ shows a declining GPP in the tropical Amazon and a stronger positive trend in central South America. After incorporating the direct CO_2 fertilization effects, both the CFE-Hybrid and CFE-ML setups predict positive trends in tropical forests, an observation absent in all other upscaled datasets. Furthermore, the CFE-Hybrid and CFE-ML models also reveal increasing GPP in temperate and boreal forests of North America and Eurasia. These patterns are also observed in BESS v2 and BEPS, while PML V2 presents minimal GPP changes in tropics and substantial reduction in Africa. Notably, all datasets agree on a pronounced GPP decrease in eastern Brazil and minimal changes in Australia.

From 2001 to 2018, a positive trend in global annual GPP is uniformly detected by all datasets, albeit with varying magnitudes (Figure 12a, Figure 13a, Figure S16~~1a-b~~). The ST_Baseline_NT model predicts a GPP growth rate of 0.35 (±0.02) Pg C year^{-2}~~per year~~, aligning with FLUXCOM-RS, but lower than FLUXSAT (0.51 Pg C year^{-2}) and MODIS (0.39 Pg C year^{-2}) ~~(Figure 11b)~~. The CFE-hybrid models estimate a notably faster GPP growth at 0.58 (±0.03) Pg C year^{-2}, similar to BESS V2 and BEPS, both around 0.55 Pg C year^{-2}. The CFE-ML models predict the highest trend~~s~~, up to 0.76 (±0.15) Pg C year^{-2} from the ST_CFE-ML_NT model and 0.59 (±0.13) Pg C year^{-2} from the ST_CFE-ML_DT model. PML V2 displays a neutral trend of 0.08 Pg C year^{-1}, and rEC-LU demonstrates an overall decline (0.20 Pg C year^{-1}). ~~Also, a higher variance is observed among ensemble members in the ST_CFE-ML setups compared to the ST_Baseline and ST_CFE-Hybrid models.~~

The LT_Baseline_NT model identifies increasing GPP trends in large areas of Europe, East and South Asia, as well as the Northern Amazon from 1982 to ~~2020~~ 2018 (Figure 11~~0b~~). The pattern from the LT_CFE-Hybrid_NT model aligns closely with the LT_Baseline_NT model but exhibit a stronger positive trend in global tropical areas as well as Eurasian boreal forests. Spatial patterns of GPP trends from BESS V2 are consistent with LT_CFE-Hybrid_NT, though with considerably higher magnitudes. ~~In contrast,~~ FLUXCOM-ERA5 shows overall negative trends in the tropics~~, with a small magnitude~~. ~~Lastly,~~ rEC-LUE agrees with CEDAR-GPP in positive GPP trends ~~identified in CEDAR-GPP~~ in the extratropical areas, but ~~predictes~~predicts a pronounced negative trend in the tropics. At the global scale, all the CEDAR-GPP long-term models predict a positive global GPP trend (Figure 12~~1bd~~, 13b). The LT_Baseline_NT and LT_Baseline_DT models show a trend of 0.13 (±0.02) and 0.15 (±0.02) Pg C year^{-2} respectively, while the LT_CFE-Hybrid_NT and LT_CFE-Hybrid_DT models double these rates with 0.33 (±0.02) and 0.31 (±0.03) Pg C year^{-2} respectively. BESS V2 predicts the highest trend at 0.61 Pg C year^{-2}. rEC-LUE shows a two-phased pattern with a strong increase in GPP from 1982 to 2000 (0.54 Pg C year^{-2}), followed by a decreasing trend after 2001 (-0.20 Pg C year^{-2}) (Figure S17~~6~~). This results in an overall positive change at a rate comparable to that of the Baseline model. FLUXCOM-ERA5 exhibited a small negative trend.



(a)
ST_Baseline_NT
ST_CFE-Hybrid_NT
ST_CFE-ML_NT
FLUXCOM-RS006
FLUXSAT
MODIS
−20
−10
0
10
20
GPP Annual Trend (gC $m^{-2} year^{-2}$)
(b)
LT_Baseline_NT
LT_CFE-Hybrid_NT
FLUXCOM-ERA5
rEC-LUE
−15
−10
−5
0
5
10
15
GPP Annual Trend (gC $m^{-2} year^{-2}$)

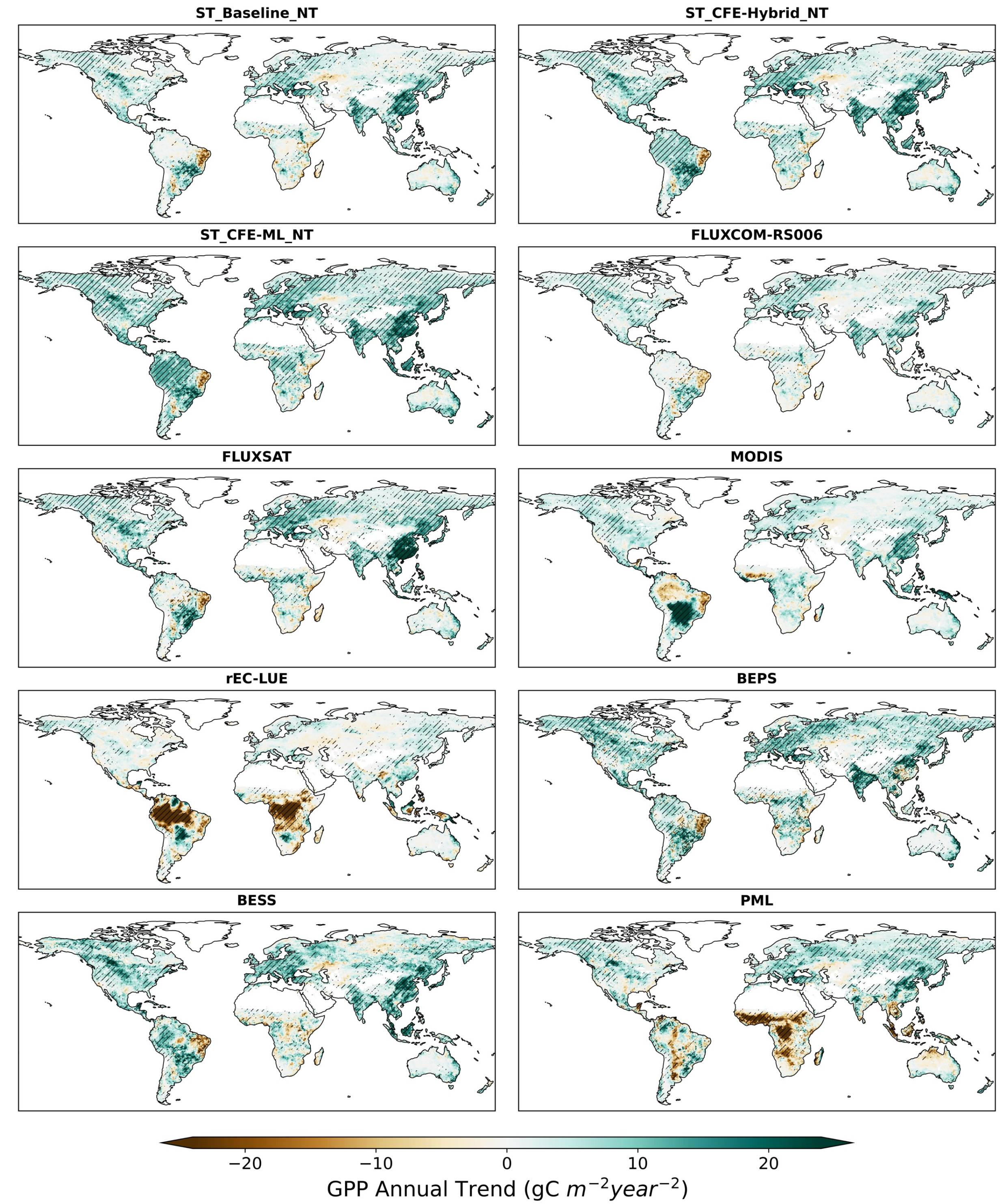


Figure 10. Annual GPP trend over 2001 – 2018 for short-term CEDAR-GPP, FLUXCOM-RS006, FLUXSAT, ~~and~~ MODIS, BESS, BEPS, and PML datasets ~~(a) and over 1982 – 2018 for long-term CEDAR-GPP, FLUXCOM-ERA5 and rEC-LUE datasets (b)~~. Hatched areas indicate the GPP trend that is statistically significant at $p < 0.05$ level under the Mann-Kendall test.

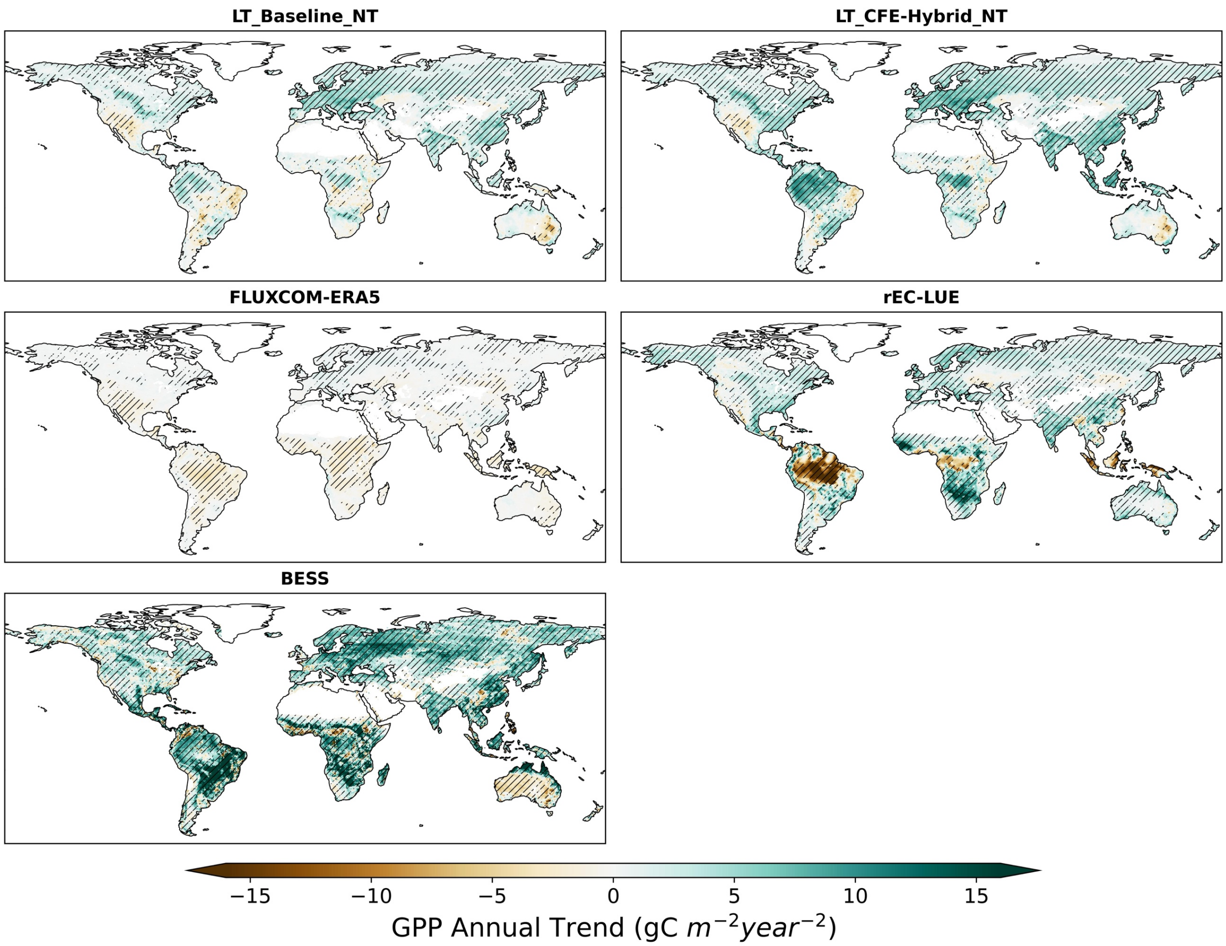
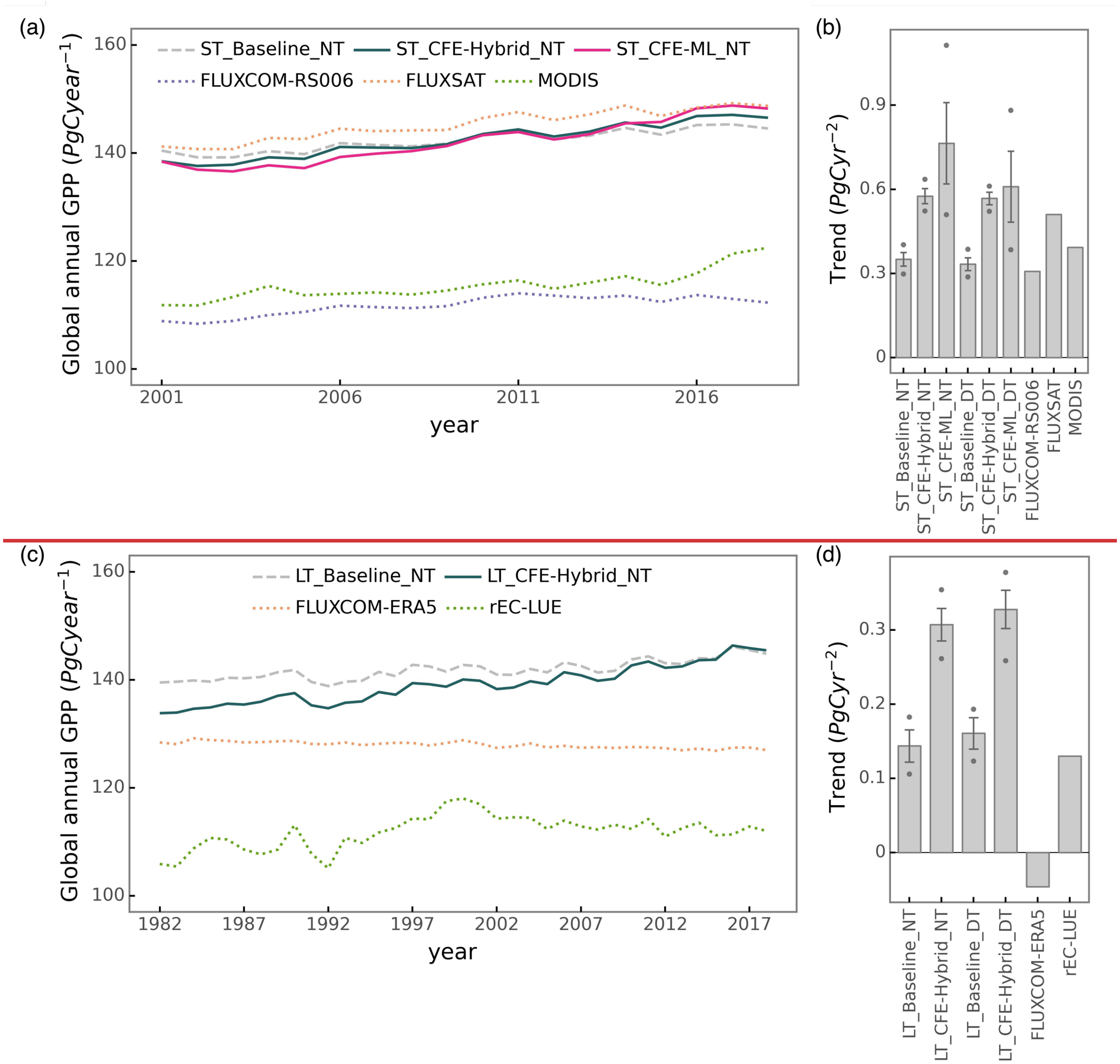


Figure 11. Annual GPP trend over 1982 – 2018 for long-term CEDAR-GPP, rEC-LUE and BESS datasets. Hatched areas indicate the GPP trend that is statistically significant at $p < 0.05$ level under the Mann-Kendall test.



(a)
ST_Baseline_NT
ST_CFE-Hybrid_NT
ST_CFE-ML_NT
FLUXCOM-RS006
FLUXSAT
MODIS
Global annual GPP ($PgCyear^{-1}$)
160
140
120
100
2001
2006
2011
2016
year
(b)
Trend ($PgCyr^{-2}$)
0.9
0.6
0.3
0
ST_Baseline_NT
ST_CFE-Hybrid_NT
ST_CFE-ML_NT
ST_Baseline_DT
ST_CFE-Hybrid_DT
ST_CFE-ML_DT
FLUXCOM-RS006
FLUXSAT
MODIS
(c)
LT_Baseline_NT
LT_CFE-Hybrid_NT
FLUXCOM-ERA5
rEC-LUE
Global annual GPP ($PgCyear^{-1}$)
160
140
120
100
1982
1987
1992
1997
2002
2007
2012
2017
year
(d)
Trend ($PgCyr^{-2}$)
0.3
0.2
0.1
0
LT_Baseline_NT
LT_CFE-Hybrid_NT
LT_Baseline_DT
LT_CFE-Hybrid_DT
FLUXCOM-ERA5
rEC-LUE

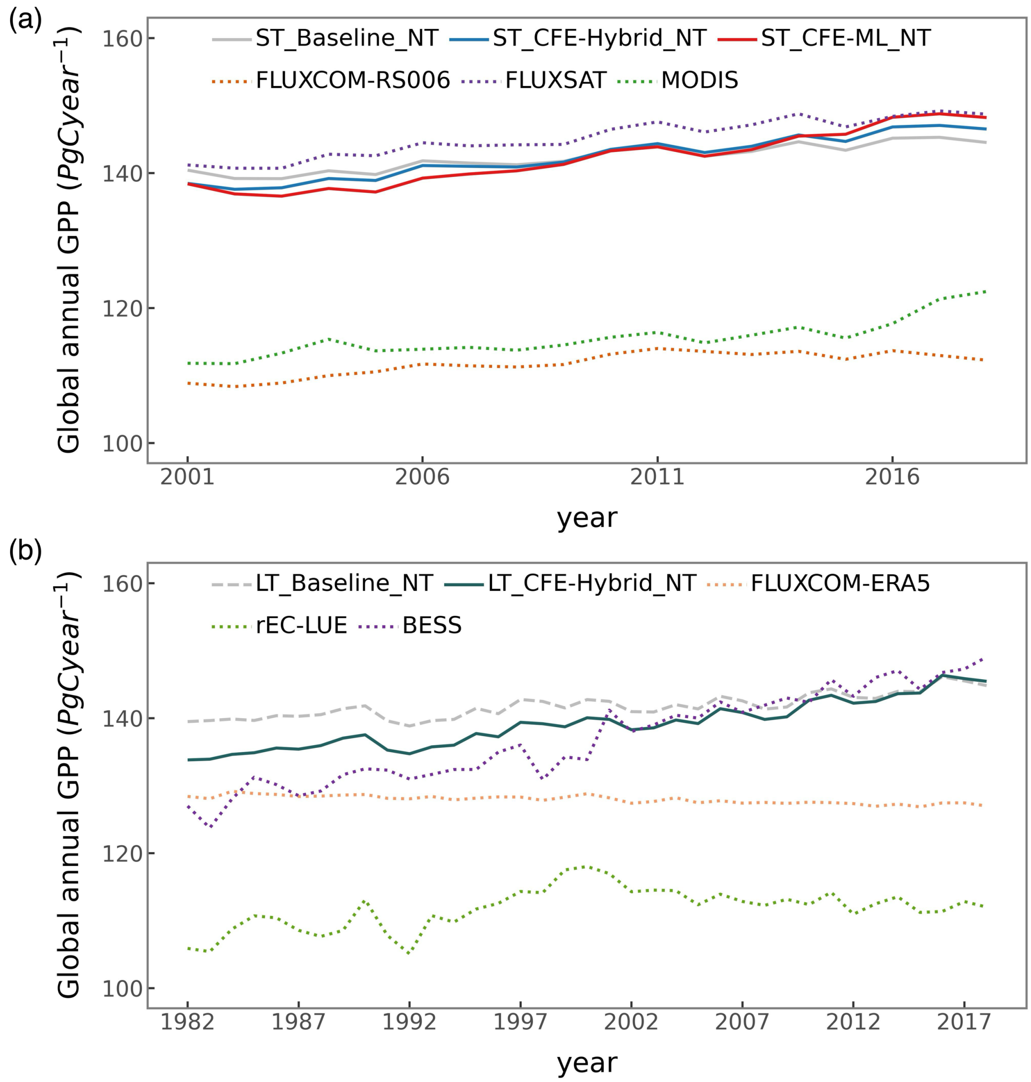


Figure 12~~1~~. Global annual GPP variations (a) ~~and trends (b)~~ from 2001 to 2018 and ~~for short-term CEDAR-GPP, FLUXCOM-RS006, FLUXSAT, and MODIS datasets. Global annual GPP variations (c) and trends~~ (b~~d~~) ~~over~~ from 1982 to 2018~~ for long-term for long-term CEDAR-GPP, FLUXCOM-ERA5, and rEC-LUE datasets~~. ~~Error bars in (b) and (d) represent the 25% to 75% percentile from the model ensembles of CEDAR-GPP. Dots in (b) and (d) indicate the minimum and maximum from the model ensembles of CEDAR-GPP.~~

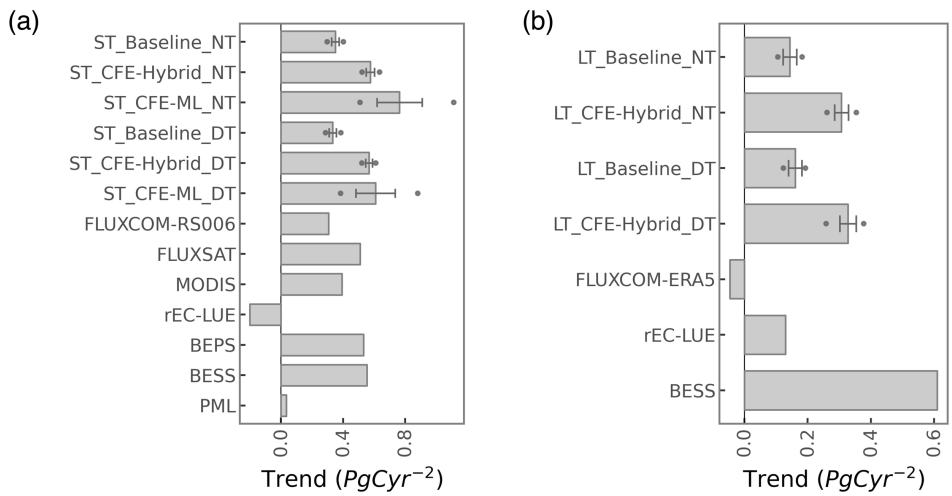


Figure 13. Global annual GPP trends for (a) 2001 to 2018 and (b) 1982 to 2018 time periods. Error bars represent the 25% to 75% percentile from the model ensembles of CEDAR-GPP. Dots indicate the minimum and maximum from the model ensembles of CEDAR-GPP.

~~Among the process-based models forced by remote sensing data, CEDAR-GPP aligns closely with BESS v2 and BEPS, showing widespread positive GPP trends in the boreal and temperate regions of the Northern Hemisphere, as well as across tropical areas (Figure S17, S18). In contrast, PML V2 presents minimal GPP changes in tropics and substantial reduction in Africa. The CEDAR-GPP CFE-ML model exhibits higher trends in tropical and northern hemisphere temperate regions, while BEPS shows more pronounced trends in boreal regions. None of the products indicates significant GPP trends in Australia. Globally, trends from BEPS and BESS V2 are around 0.55 PgC year^{-1}, consistent with CEDAR's ST_CFE-Hybrid_NT model at 0.58 PgC year^{-1}, while PML V2 displays a neutral trend of 0.08 PgC year^{-1}.~~

3.3 GPP estimation uncertainties

We analyzed the spread between the 30 model ensemble members in CEDAR-GPP as an indicator of uncertainties in GPP estimations. The spatial pattern of uncertainty in estimating annual mean GPP largely resembles that of the mean map (Figure 14~~2~~, Figure 6a). The largest model spread is found in highly productive tropical forests, and this uncertainty decrease in temperate and cold areas (Figure 14~~2~~a). Tropical ecosystems, with a mean annual GPP between 1000 to 3500 Pg C year^{-1}, only exhibit a 2% and 6% variation within the model ensemble (Figure 14~~2~~b). Ecosystems in the temperate and cold climates have a smaller annual GPP and proportionally small uncertainties of up to 6%.

However, ecosystems in Arid and Polar climates, despite their similarly low GPP, show higher model uncertainty, reaching 10% to 40% of the ensemble mean.

The estimation uncertainty of GPP trends is generally below 15% to 20% in the CEDAR-GPP datasets under the ST_Baseline and ST_CFE-Hybrid setups (Figure ~~12c~~14c). However, in the ST_CFE-ML setup, the estimation increases substantially, with model spread reaching up to 40% in tropical areas. Figure 15 (Figure S19) ~~S20~~ further illustrates the trend uncertainties with the ensemble mean error range based on one standard deviation. Both the CFE-ML models show large discrepancies between the upper and lower uncertainty ranges particularly within the tropics. Additionally, the long-term models also show a higher uncertainty compared to the short-term models.

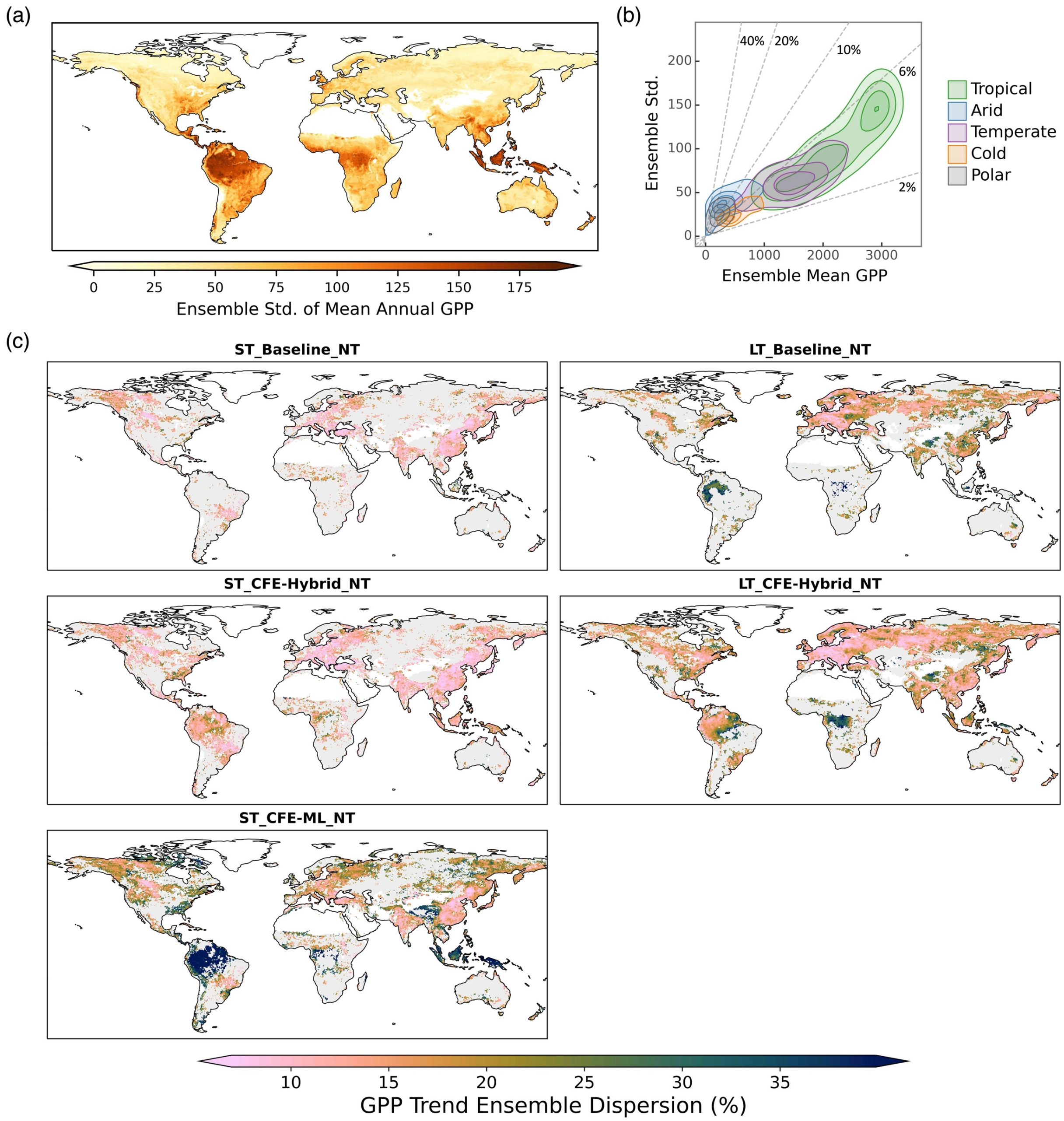


Figure 14~~2~~. CEDAR-GPP estimation uncertainty derived from ensemble spread (standard deviation of 30 model predictions). (a) Spatial patterns of the absolute standard deviation from ensemble members in estimating the mean annual GPP from 2001 to 2018, using data from the ST_CFE-Hybrid_NT setup. (b) Relationships between ensemble standard deviation and ensemble mean in mean annual GPP. Colored contours denote clusters of Koppen climate zones. Dashed lines indicate the ratio between the ensemble standard deviation and the ensemble mean with values shown in percentage. (c) Spatial patterns of model uncertainty in GPP long-term trend estimation. Only areas where 90% of the ensemble members showed a statistically significant trend ($p<0.05$) are shown in the maps. The trend for the short-term datasets (left column) was computed between 2001 to 2018. The trend for the long-term datasets (right column) was computed between 1982 to 2018.

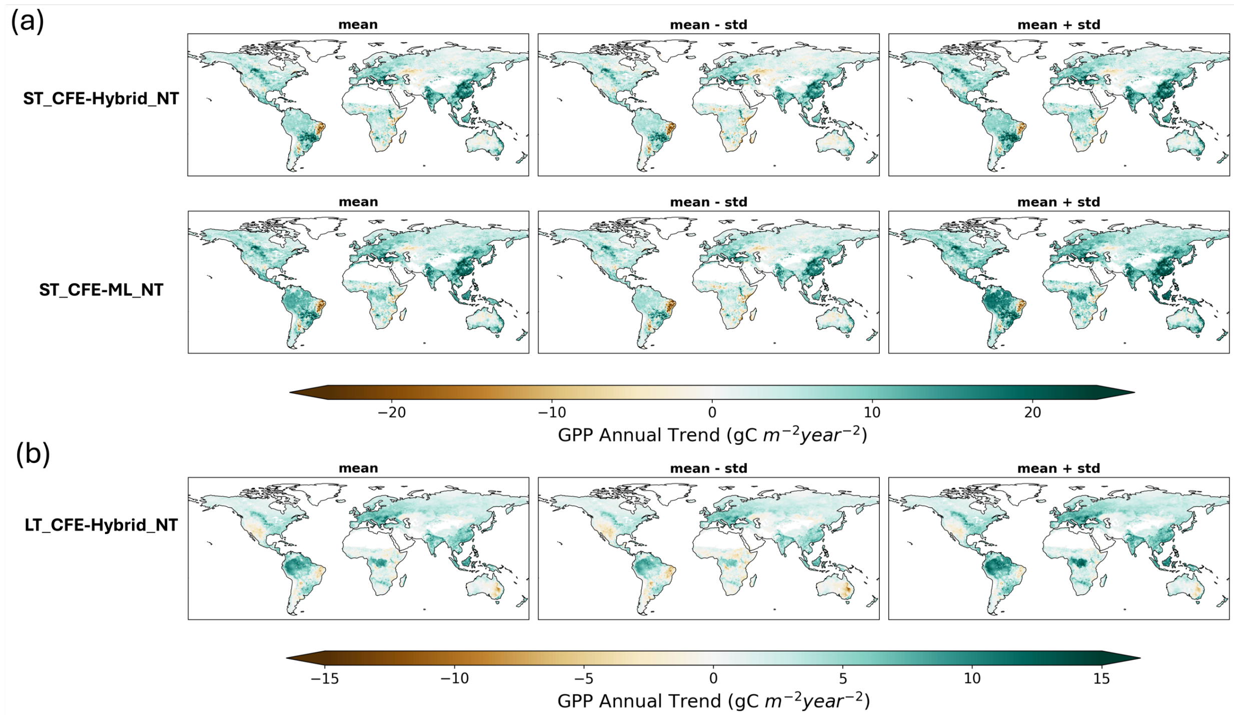


Figure 15. Maps of GPP trends and uncertainty range for CEDAR-GPP CFE datasets (NT only). The first column presents ensemble mean trends, the second column shows trends from the mean minus one standard deviation (upper, and third column indicates the trend from the mean plus one standard deviation. (a) Trends from the short-term (ST) datasets evaluated from 2001 to 2020. (b) Trends from the long-term (LT) dataset evaluated from 1982 to 2020. DT datasets were shown in Figure S19.

4. Discussion

4.1 Reducing uncertainties in GPP upscaling

Here we examine the three predominant sources of uncertainties in machine learning upscaling of GPP: eddy covariance measurements, input datasets, and the machine learning model. We discuss strategies used in CEDAR-GPP to reduce the impacts of these uncertainties and highlight potential future research directions.

4.1.1 Eddy covariance data

Uncertainties associated with eddy covariance measurement and data processing can propagate through the upscaling process. CEDAR-GPP was produced using monthly aggregated eddy covariance data, where the impact of random errors in half-hourly measurements was minimized due to the temporal aggregation (Jung et al., 2020). Our stringent quality screening further reduced data processing uncertainties such as those associated with gap-filling. Yet, the discrepancy in GPP patterns between the CEDAR-GPP NT and DT setups is indicative of systematic biases linked to the partitioning approaches used to derive GPP from the NEE measurements (Keenan et al., 2019; Pastorello et al., 2020). Interestingly, the mean annual GPP from the DT setup is slightly higher than that from the NT setup (Figure 6), and the DT setup also predicts a higher GPP trend in the long-term dataset (Figure ~~11~~13). While these discrepancies are relatively small compared to the predominant spatiotemporal patterns, the separate DT and NT setups in CEDAR-GPP offer an interesting quantification of the GPP partitioning uncertainties over space and time, providing insights for future methodology improvements.

The unbalanced spatial representativeness of the eddy covariance data constitutes a more significant source of uncertainty, as highlighted by previous studies (Jung et al., 2020; Tramontana et al., 2015). Effective generalization of machine learning models requires a substantial volume of training data that adequately represents and balances varied conditions. In CEDAR-GPP, this issue was mitigated with a large set of eddy covariance data (~18000 site-months) integrating FLUXNET2015 and two regional networks. However, data availability remains limited in critical carbon exchange hotspots such as tropics, subtropics, drylands, and boreal regions, as well as in mountainous areas (Figure 1). Contrary to widespread perception that sparse training data leads to high upscaling uncertainties, our findings from the bootstrapped model spread indicates modest uncertainties in tropical areas relative to their high GPP magnitude (Figure ~~12~~14). This observation aligns with findings from the FLUXCOM product, revealing low extrapolation uncertainty in humid tropical regions (Jung et al., 2020). Nevertheless, to fully understand the upscaling uncertainty, it is essential to evaluate the generalization or extrapolation errors within the predictor space and consider the potential limitations of model structures (van der Horst et al., 2019; Villarreal and Vargas, 2021). Additionally, data limitations in mountainous areas and the absence of topology information in the predictor space in our models suggest potential uncertainties related to topographical effects on GPP (Hao et al., 2022; Xie et al., 2023).

Furthermore, our analysis suggests that the estimated global GPP magnitudes ~~were~~ are related to the specific eddy covariance GPP data used in upscaling. Notably, global GPP magnitudes derived from CEDAR-GPP closely align with those from FLUXSAT, while the estimates from FLUXCOM were considerably lower (Figure 6, Figure ~~11~~12). FLUXSAT used eddy covariance data from FLUXNET2015, which largely overlapped with that included in CEDAR-GPP (Joiner and Yoshida, 2020). FLUXCOM utilized data from FLUXNET La Thuile set and CarboAfrica network, which consists of a distinct set of sites (Tramontana et al., 2016). The influence from the predictor datasets is minimal since all three datasets relied on MODIS-derived products. For a more in-depth evaluation of the impacts of flux site representativeness on upscaling, future research directions could include conducting synthetic experiments with simulations of ensembles of terrestrial biosphere models.

4.1.2 Input predictors and controlling factors

Upscaled GPP inherent uncertainties from the input predictors, including satellite and climate datasets. First, satellite remote sensing data contains noises resulting from sun-earth geometry, atmospheric conditions, soil background, and geolocation inaccuracies. The models or algorithms used for retrieving LAI, fAPAR, LST, and soil moisture, also contain random errors and systematic biases specific to certain regions, biome types, or climatic conditions (Fang et al., 2019; Ma et al., 2019; Yan et al., 2016b). Moreover, satellite observations frequently contain missing values due to clouds, aerosols, snow, and algorithm failure, leading to both systematic and random uncertainties. In producing CEDAR-GPP, we mitigated these uncertainties through comprehensive preprocessing procedures. Our temporal gap-filling strategy exploits both the temporal dependency of vegetation status and long-term climatology, to reduce biases from missing values. Temporal and spatial aggregation further ~~reduced~~ reduces the remaining data gaps and random noises. Nevertheless, considerable uncertainties likely remain in satellite datasets impacting the upscaled estimations.

A potentially more impactful source of uncertainty is the mismatch between the footprint of the eddy covariance measurements and the coarse resolution of satellite observations. While flux towers typically have a footprint of around ~1 km^2 (Chu et al., 2021), satellite observations employed in CEDAR-GPP and most other upscaled datasets are at 5 km or lower resolution. Systematic and random errors could be introduced due to this mismatch, particularly in heterogenous biomes and areas with a mixture of vegetation and non-vegetated land covers. One mitigation strategy is to generate upscaled datasets at a higher spatial resolution (e.g. 500m). Alternatively, models could be trained at a high resolution and applied to the coarse resolution to reduce computation and storage requirements (Dannenberg et al., 2023; Gaber et al., 2024). However, this approach does not address inherent scaling errors in coarse-resolution satellite images (Dong et al., 2023; Yan et al., 2016a).

Besides the quality of predictors, successful machine learning upscaling also requires a comprehensive set of features representing all controlling factors. For example, the lack of GPP interannual variabilities in FLUXCOM-ERA5 manifests the importance of incorporating dynamic vegetation signals from remote sensing in the upscaling framework. CEDAR-GPP used satellite observations from optical, thermal, and microwave systems as well as climate variables thoroughly representing GPP dynamics. Particularly, the inclusion of LST and soil moisture data provides important information about resource limitations and stress factors, which are crucial for certain biomes and/or under specific conditions (Green et al., 2022; Stocker et al., 2018, 2019). Dannenberg et al. (2023) showed that incorporating LST from MODIS and soil moisture from the SMAP satellite datasets substantially improved the machine learning estimation accuracy of GPP in North American drylands. Nevertheless, accurately capturing interannual anomalies remains challenging for certain biomes, such as evergreen needleleaf forest, cropland, and wetland (Figure 4), as acknowledged by previous studies (Tramontana et al., 2016; Jung et al., 2020). High prediction uncertainties (Figure 14, 15~~2~~) in drylands also suggest the machine learning models did not sufficiently represent the mechanisms of water stress and drought responses. Potential improvement may be achieved by incorporating datasets related to agricultural management practices (crop type, cultivar, irrigation, fertilization) (Xie et al., 2021), plant hydraulic and physiological properties (Liu et al., 2021), dynamic C4 plant distributions (Luo et al., 2024), root and soil characteristics (Stocker et al., 2023), as well as topography (Xie et al., 2023).

4.1.3 Machine learning models and uncertainty quantification

The choice of machine learning models and their parameterization has been found to have a relatively minor impact on GPP upscaling uncertainties (Tramontana et al., 2015). CEDAR used the state-of-the-art boosting algorithm, XGBoost, which provided high performance given the current data availability. Further reduction of model uncertainty will likely rely on additional information, such as increasing the number of eddy covariance sites or incorporating more high-quality predictors. Additionally, temporal dependency of carbon fluxes responses to atmospheric controls may also be exploited with specialized deep neural networks such as recurrent neural networks or transformers (Besnard et al., 2019; Ma and Liang, 2022).

A key challenge, however, is the quantification of uncertainties in machine learning upscaling (Reichstein et al., 2019). The limited availability of eddy covariance data hinders a comprehensive assessment of the extrapolation errors; consequently, metrics of predictive performance from cross-validation are inherently biased. CEDAR derived estimation uncertainty for each GPP prediction using bootstrapping model ensemble, which naturally mimics the sampling bias associated with flux tower locations. Notably, the choice of input climate reanalysis datasets could also induce systematic differences in GPP spatial and temporal patterns (Tramontana et al., 2015). As a result, the FLUXCOM product generates model ensembles based on different reanalysis datasets to capture these uncertainties. Additionally, different satellite datasets of vegetation structural proxies, such as LAI, also exhibit~~ed~~ significant discrepancies (Jiang et al., 2017). Thus, an ensemble approach combining site-level bootstrapping with multiple sources of input predictors could potentially provide a more comprehensive quantification of uncertainties. Furthermore, tree-based models do not generalize well to unseen conditions, and the uncertainty estimates derived from bootstrapping of XGBoost models may underrepresent actual biases stemming from limitations in training data representation. Future work may explore Bayesian neural networks, which provide uncertainty along with predictions and, at the same time, present high predictive power comparable to ensemble tree-based algorithms (Ma et al., 2021).

4.2 Long-term GPP changes and CO_2 fertilization effect

CEDAR-GPP was constructed using a comprehensive set of climate variables and multi-source satellite observations, thus encapsulating long-term GPP dynamics from both direct and indirect effects of climate controls. Particularly, CEDAR-GPP included the direct CO_2 fertilization effect, which has been shown to dominate the increasing trend of global photosynthesis (Chen et al., 2022). Incorporating these effects substantially improved long-term trends of GPP from site to global scales (Figure 5, 10, 11, 12, 13). CEDAR's CFE-Hybrid setup offeres a conservative estimation of the direct CO_2 effects by simulating the CO_2 sensitivity of light-limited LUE for C3 plants (Walker et al., 2021). However, the model does not account for the impacts of nutrient availability, which could potentially constrain CO_2 fertilization (Peñuelas et al., 2017; Reich et al., 2014; Terrer et al., 2019). Robust modeling of LUE responses to rising CO_2 under various environmental conditions remains challenging (Wang et al., 2017). Future work is needed to better understand how these factors affect the quantification of GPP and its long-term temporal variations.

The CFE-ML model adopted a data-driven approach to infer CO_2 effects directly from eddy covariance data. This strategy allows the model to potentially capture multiple physiological pathways of the CO_2 impact evidenced in the eddy covariance measurements, including the increases of the biochemical rates and enhancements in the water use efficiency (Keenan et al., 2013). The model detects a strong positive effect of CO_2 on eddy covariance measured GPP, consistent with previous studies based on process-based and statistical models (Chen et al., 2022; Fernández-Martínez et al., 2017; Ueyama et al., 2020). Moreover, spatial patterns of GPP trends derived from the CFE-ML model reflected a strong temperature dependency, aligning with the anticipated temperature sensitivity of photosynthetic biochemical processes (Keenan et al., 2023). Yet, the considerable ensemble spread in the CO_2 trends from the CFE-ML model and discrepancies between the CFE setups (Figure 14~~1, Figure 13~~) underscores a high level of uncertainty in the machine learning quantified CO_2 effects.

Several limitations should be noted regarding GPP trend estimation and validation. First, the CFE-ML model may not fully capture the intricate mechanisms of plant physiological responses to CO_2. For example, eddy covariance towers, especially long-term sites, are typically located in homogeneous and undisturbed ecosystems, not representative of the full diversity of ecosystems globally. Thus, interactions between CO_2 and natural or human-induced disturbance, as well as many other stresses, are likely underrepresented in the models. Ultimately, the model's capacity to robustly quantify CO_2 fertilization is constrained by the scope and diversity of the eddy covariance data. Additionally, the use of spatially invariant CO_2 data may not fully represent the actual CO_2 variations that plants experience across different environments.

Secondly, CO_2 effects inferred by the CFE-ML models may be confounded by other factors that correlate with CO_2 over time. Industrialization-induced nitrogen deposition could synergistically boost GPP alongside CO_2 (O'Sullivan et al., 2019). Technological and management improvements in agriculture that contribute to a global enhancement of crop photosynthesis (Zeng et al., 2014), might also be indirectly reflected in the model estimates. Moreover, interactions with the other input features that exhibit long-term trends, such as those induced by non-biological factors (e.g. sensor orbital drifts), also affect the CO_2 effects inference. Additionally, other factors that could lead to long-term GPP trends (e.g. forest aging, disturbances) might also be underrepresented in our models.

Finally, direct validation of GPP trends is limited, particularly in tropical regions, constrained by the availability of long-term records. Detecting and evaluating trends is challenging and typically requires long monitoring records (e.g. over 10 to 15 years), since long-term changes, such as those induced by CO_2, are very small relative to large interannual variations. Evaluating aggregated GPP trends across multiple sites presents an alternative approach; however, there were still insufficient sites in tropical and evergreen broadleaf forest areas to robustly validate our estimates for those ecosystems (Figure 5). Partly due to data limitations, uncertainties in GPP estimated from bootstrapped samples are very high in tropical areas (Figure ~~12~~14). Thus, trend estimates in these areas should be interpreted in the context of associated uncertainties and limitations.

Our results also suggested that variations in the estimated GPP long-term trends from different products are largely related to the representation of CO_2 fertilization. Products that do not consider the direct CO_2 effect, including our Baseline models, FLUXSAT, FLUXCOM, and MODIS, show minimal long-term changes in tropical GPP, while the CEDAR CFE-ML and CFE-Hybrid models demonstrate significant GPP increases aligning with predictions from the terrestrial biosphere models (Anav et al., 2015). FLUXCOM-ERA5, not accounting for dynamics changes in vegetation structures and CO_2, does not capture either the direct or indirect CO_2 fertilization resulting in a slight negative GPP trend attributable to shifted climate patterns. Notably, rEC-LUE exhibit contrasting trends before and after circa 2000, primarily attributed to changes in vapor pressure deficit, PAR, and LAI, while the direct CO_2 fertilization effect remains consistent (Zheng et al., 2020). CEDAR CFE-ML and CFE-Hybrid models align well with two process-based models forced with remote sensing data which consider direct CO_2 effects (BESS and BEPS). Nevertheless, considerable differences between CEDAR-GPP and other remote sensing products that include direct CO_2 effects (rEC-LUE and PML V2) warrant more in-depth investigations into long-term GPP responses to changes in atmospheric CO_2 and climate patterns.

Lastly, quantifications of GPP trends and their causes remain highly uncertain from site to global scales. Trend detection is often complicated by data noises and interannual variabilities, thus requiring long-term records which are limited in certain areas, biomes, and environmental conditions, such as tropics, polar regions, wetlands, as well as ecosystems with regular or anthropogenic disturbances (Baldocchi et al., 2018; Zhan et al., 2022). Moreover, isolating the effect of CO_2 is challenging, as it is confounded by other factors, such as forest regrowth, land cover change, and disturbances, which also significantly impacts long-term GPP variations. To this end, continued efforts in expanding ecosystem flux measurements and standardizing data processing present new opportunities to assess ecosystem productivity responses to changing climate conditions (Delwiche et al., 2024; Pastorello et al., 2020). Future research could also leverage novel machine learning techniques, such as knowledge-guided machine learning (Liu et al., 2024) and hybrid modeling that combines process-based and machine learning approaches (Kraft et al., 2022; Reichstein et al., 2019).

5. Data availability and usage note

The CEDAR-GPP product, comprising ten GPP datasets, can be accessed at https://zenodo.org/doi/10.5281/zenodo.8212706 (Kang et al., 2024). These datasets were generated at a spatial resolution of 0.05° and monthly time steps. Each dataset includes an ensemble mean GPP (“GPP_mean”) and an ensemble standard deviation (“GPP_std”). Data is formatted in netCDF with the following naming convention: “CEDAR-GPP_<version>_<model setup>_<YYYYMM>.nc”.

The CEDAR GPP product offers GPP estimates derived from ten different models. Models are characterized by 1) temporal coverage, 2) configuration of CO_2 fertilization, and 3) GPP partitioning approach (Table 2). We provide a structured approach to selecting the most appropriate dataset for research or applications.

1) Study period considerations: the Short-Term (ST) setup is ideal for studies focusing on periods after 2000. These models are constructed using a broader range of explanatory predictors, offering higher precision and smaller random errors. The Long-Term (LT) datasets shall be used for research assessing GPP dynamics over a longer time period (before 2001). It is important to note that trends from the ST and LT datasets are not directly comparable, as they were derived from different satellite remote sensing data.

2) CO_2 Fertilization Effect (CFE) configurations: the CFE-Hybrid and CFE-ML setups are preferable when assessing temporal GPP dynamics, especially long-term trends. The CFE-Hybrid setup includes a hypothetical trend from the direct CO_2 effect, while CFE-ML is purely data-driven and does not make any specific assumption about the sensitivity of photosynthesis to CO_2. Averaging the CFE-Hybrid and CFE-ML estimates is acceptable, with the difference between them reflecting the uncertainty surrounding the direct CO_2 effect. Note that the Baseline setup should not be used to study long-term GPP dynamics, especially those induced by elevated CO_2. The Baseline setup may be useful to compare with other remote sensing-derived GPP datasets that do not consider the direct CO_2 effect. Differences between these setups regarding mean GPP spatial patterns, seasonal and interannual variations are considered to be minor.

3) GPP partitioning methods: We recommend using the mean value derived from both the “NT” (Nighttime) and “DT” (Daytime). The difference between these two provides insight into the uncertainties arising from the partitioning approaches used in GPP estimation from eddy covariance measurements.

Finally, like other upscaled or remote sensing-based GPP datasets, CEDAR-GPP should not be regarded as "observations" but rather as model estimates informed by remote sensing and ground-based data. The extent of assumptions or structural constraints varies across such datasets. CEDAR-GPP, particularly in its CFE-Baseline and CFE-ML configurations, is entirely data-driven and incorporates no explicit assumptions regarding the biological and environmental processes underlying photosynthesis, apart from the generic assumptions inherent in machine learning models. Consequently, the usage and interpretation of this dataset should be carefully framed within the context of the input eddy covariance and environmental data as well as their limitations.

6. Code availability

The code for upscaling and generating global GPP datasets can be accessed at https://doi.org/10.5281/zenodo.8400968.

7. Conclusions

We present the CEDAR-GPP product generated by upscaling global eddy covariance measurements with machine learning and a broad range of satellite and climate variables. CEDAR-GPP comprises four long-term datasets from 1982 to 2020 and six short-term datasets from 2001 to 2020. These datasets encompass three configurations regarding the incorporation of direct CO_2 fertilization effects and two partitioning approaches to derive GPP from eddy covariance data. The machine learning models of CEDAR-GPP demonstrated high capability in predicting monthly GPP, its seasonal cycles, and spatial variability within the global eddy covariance sites, with cross-validated R^2 between 0.56 to 0.79. Short-term model setups consistently outperformed long-term models due to considerably more and higher-quality information from multi-source satellite observations.

CEDAR-GPP advances satellite-based GPP estimations, as the first upscaled dataset that considered the direct biochemical effects of elevated atmospheric CO_2 on photosynthesis, which is responsible for an increasing land carbon sink over the past decades. We show that incorporating this effect in our CFE-ML and CFE-Hybrid models substantially improved the estimation of GPP trends at eddy covariance sites. Global patterns of long-term GPP trends in the CFE-ML setups show a strong temperature dependency consistent with biophysical theories. However, trend estimation and validation remain particularly challenging in data-scarce regions, such as the tropics, emphasizing the need for enhanced data availability and methodological advancements. Beyond trends, global spatial and temporal GPP patterns from CEDAR generally align with other satellite-based GPP datasets.

In conclusion, CEDAR-GPP, informed by global eddy covariance measurements and a broad range of multi-source remote sensing observations and climatic variables, offers a comprehensive representation of global GPP spatial and temporal dynamics over the past four decades. The different CO_2 fertilization configurations integrated in CEDAR-GPP offer new opportunities for understanding global ecosystem photosynthesis's response to increases in atmospheric CO_2 along different pathways over space and time. CEDAR-GPP is expected to serve as a valuable tool for benchmarking process-based modeling and constraining the global carbon cycle.

Appendix A: List of eddy covariance sites

Site ID	IGBP	Data Range	Citation
AR-SLu	MF	2010 - 2011	(Garcia et al., 2016)
AR-Vir	ENF	2010 - 2012	(Posse et al., 2016)
AT-Neu	GRA	2002 - 2012	(Wohlfahrt et al., 2016)
AU-Ade	SAV	2010 - 2014	(Beringer and Hutley, 2016a)
AU-ASM	WSA	2007 - 2009	(Cleverly et al., 2016)
AU-Cpr	SAV	2010 - 2014	(Meyer et al., 2016)
AU-Cum	EBF	2012 - 2014	(Pendall et al., 2016)
AU-DaP	GRA	2007 - 2013	(Beringer and Hutley, 2016b)
AU-DaS	SAV	2008 - 2014	(Beringer and Hutley, 2016g)
AU-Dry	SAV	2008 - 2014	(Beringer and Hutley, 2016c)
AU-Emr	GRA	2011 - 2013	(Schroder et al., 2016)
AU-Fog	WET	2006 - 2008	(Beringer and Hutley, 2016d)
AU-Gin	WSA	2011 - 2014	(Macfarlane et al., 2016)
AU-How	WSA	2001 - 2014	(Beringer and Hutley, 2016e)
AU-RDF	WSA	2011 - 2013	(Beringer and Hutley, 2016f)
AU-Rig	GRA	2011 - 2014	(Beringer et al., 2016a)
AU-Tum	EBF	2001 - 2014	(Woodgate et al., 2016)
AU-Wac	EBF	2005 - 2008	(Beringer et al., 2016b)
AU-Whr	EBF	2011 - 2014	(Beringer et al., 2016c)
AU-Wom	EBF	2010 - 2014	(Arndt et al., 2016)
AU-Ync	GRA	2012 - 2014	(Beringer and Walker, 2016)
BE-Bra	MF	2001 - 2020	(Warm Winter 2020 Team, 2022)
BE-Dor	GRA	2011 - 2020	(Warm Winter 2020 Team, 2022)
BE-Lon	CRO	2004 - 2020	(Warm Winter 2020 Team, 2022)
BE-Maa	CSH	2016 - 2020	(Warm Winter 2020 Team, 2022)
BE-Vie	MF	2001 - 2020	(Warm Winter 2020 Team, 2022)
BR-Sa1	EBF	2002 - 2011	(Saleska, 2016)
BR-Sa3	EBF	2001 - 2004	(Goulden, 2016a)
CA-Ca1	ENF	2001 - 2002	(Black, 2023a)
CA-Ca2	ENF	2001 - 2010	(Black, 2023b)
CA-Ca3	ENF	2001 - 2010	(Black, 2018)
CA-Cbo	DBF	2001 - 2003	(Staebler, 2022)
CA-Gro	MF	2003 - 2014	(McCaughey, 2022)
CA-Man	ENF	2001 - 2008	(Amiro, 2016a)
CA-NS1	ENF	2002 - 2005	(Goulden, 2022a)
CA-NS2	ENF	2001 - 2005	(Goulden, 2022b)
CA-NS3	ENF	2001 - 2005	(Goulden, 2022c)
CA-NS4	ENF	2002 - 2005	(Goulden, 2016b)
CA-NS5	ENF	2001 - 2005	(Goulden, 2022d)
CA-NS6	OSH	2001 - 2005	(Goulden, 2022e)
CA-NS7	OSH	2002 - 2005	(Goulden, 2016c)
CA-Oas	DBF	2001 - 2010	(Black, 2016a)
CA-Obs	ENF	2001 - 2010	(Black, 2016b)
CA-Qc2	MF	2008 - 2010	(Margolis, 2018)
CA-Qfo	ENF	2003 - 2010	(Margolis, 2023)
CA-SF1	ENF	2003 - 2006	(Amiro, 2016b)
CA-SF2	ENF	2003 - 2005	(Amiro, 2023)
CA-SF3	OSH	2003 - 2006	(Amiro, 2016c)

CA-SJ2	ENF	2003 - 2007	(Barr and Black, 2018)
CA-TP1	ENF	2003 - 2014	(Arain, 2016b)
CA-TP2	ENF	2003 - 2007	(Arain, 2016c)
CA-TP3	ENF	2003 - 2014	(Arain, 2016d)
CA-TP4	ENF	2003 - 2017	(Arain, 2016a)
CA-TPD	DBF	2012 - 2014	(Arain, 2016e)
CA-WP1	WET	2003 - 2009	(Flanagan, 2018a)
CA-WP2	WET	2004 - 2006	(Flanagan, 2018b)
CA-WP3	WET	2004 - 2006	(Flanagan, 2018c)
CG-Tch	SAV	2006 - 2009	(Nouvellon, 2016)
CH-Aws	GRA	2006 - 2020	(Warm Winter 2020 Team, 2022)
CH-Cha	GRA	2005 - 2020	(Warm Winter 2020 Team, 2022)
CH-Dav	ENF	2001 - 2020	(Warm Winter 2020 Team, 2022)
CH-Fru	GRA	2005 - 2020	(Warm Winter 2020 Team, 2022)
CH-Lae	MF	2004 - 2020	(Warm Winter 2020 Team, 2022)
CH-Oe1	GRA	2002 - 2008	(Ammann, 2016)
CH-Oe2	CRO	2004 - 2020	(Warm Winter 2020 Team, 2022)
CN-Cha	MF	2003 - 2005	(Zhang and Han, 2016)
CN-Cng	GRA	2007 - 2010	(Dong, 2016)
CN-Din	EBF	2003 - 2005	(Zhou and Yan, 2016)
CN-Du2	GRA	2007 - 2008	(Chen, 2016c)
CN-Ha2	WET	2003 - 2005	(Li, 2016)
CN-HaM	GRA	2002 - 2004	(Tang et al., 2016)
CN-Qia	ENF	2003 - 2005	(Wang and Fu, 2016)
CN-Sw2	GRA	2011 - 2012	(Shao, 2016)
CZ-BK1	ENF	2004 - 2020	(Warm Winter 2020 Team, 2022)
CZ-BK2	GRA	2006 - 2012	(Sigut et al., 2016)
CZ-KrP	CRO	2014 - 2020	(Warm Winter 2020 Team, 2022)
CZ-Lnz	DBF	2015 - 2020	(Warm Winter 2020 Team, 2022)
CZ-RAJ	ENF	2012 - 2020	(Warm Winter 2020 Team, 2022)
CZ-Stn	DBF	2010 - 2020	(Warm Winter 2020 Team, 2022)
CZ-wet	WET	2006 - 2020	(Warm Winter 2020 Team, 2022)
DE-Akm	WET	2009 - 2020	(Warm Winter 2020 Team, 2022)
DE-Geb	CRO	2001 - 2020	(Warm Winter 2020 Team, 2022)
DE-Gri	GRA	2004 - 2020	(Warm Winter 2020 Team, 2022)
DE-Hai	DBF	2001 - 2020	(Warm Winter 2020 Team, 2022)
DE-HoH	DBF	2015 - 2020	(Warm Winter 2020 Team, 2022)
DE-Hte	WET	2009 - 2018	(Drought 2018 Team, 2020)
DE-Hzd	DBF	2010 - 2020	(Warm Winter 2020 Team, 2022)
DE-Kli	CRO	2004 - 2020	(Warm Winter 2020 Team, 2022)
DE-Lkb	ENF	2009 - 2013	(Lindauer et al., 2016)
DE-Lnf	DBF	2002 - 2012	(Knohl et al., 2016)
DE-Obe	ENF	2008 - 2020	(Warm Winter 2020 Team, 2022)
DE-RuR	GRA	2011 - 2020	(Warm Winter 2020 Team, 2022)
DE-RuS	CRO	2011 - 2020	(Warm Winter 2020 Team, 2022)
DE-RuW	ENF	2012 - 2020	(Warm Winter 2020 Team, 2022)
DE-Seh	CRO	2007 - 2010	(Schneider and Schmidt, 2016)
DE-SfN	WET	2012 - 2014	(Klatt et al., 2016)
DE-Spw	WET	2010 - 2014	(Bernhofer et al., 2016)
DE-Tha	ENF	2001 - 2020	(Warm Winter 2020 Team, 2022)
DK-Eng	GRA	2005 - 2007	(Pilegaard and Ibrom, 2016)
DK-Sor	DBF	2001 - 2020	(Warm Winter 2020 Team, 2022)

ES-Abr	WSA	2015 - 2020	(Warm Winter 2020 Team, 2022)
ES-Agu	OSH	2006 - 2019	(Warm Winter 2020 Team, 2022)
ES-Amo	OSH	2007 - 2012	(Poveda et al., 2016)
ES-LgS	OSH	2005 - 2020	(Reverter et al., 2016)
ES-LJu	WSA	2014 - 2020	(Warm Winter 2020 Team, 2022)
ES-LM1	WSA	2014 - 2020	(Warm Winter 2020 Team, 2022)
ES-LM2	OSH	2007 - 2009	(Warm Winter 2020 Team, 2022)
FI-Hyy	ENF	2001 - 2020	(Warm Winter 2020 Team, 2022)
FI-Jok	CRO	2001 - 2003	(Lohila et al., 2016)
FI-Ken	ENF	2018 - 2020	(Warm Winter 2020 Team, 2022)
FI-Let	ENF	2009 - 2020	(Warm Winter 2020 Team, 2022)
FI-Lom	WET	2007 - 2009	(Aurela et al., 2016a)
FI-Qvd	CRO	2018 - 2020	(Warm Winter 2020 Team, 2022)
FI-Sii	GRA	2016 - 2020	(Warm Winter 2020 Team, 2022)
FI-Sod	ENF	2001 - 2014	(Aurela et al., 2016b)
FI-Var	ENF	2016 - 2020	(Warm Winter 2020 Team, 2022)
FR-Aur	CRO	2005 - 2020	(Warm Winter 2020 Team, 2022)
FR-Bil	ENF	2014 - 2020	(Warm Winter 2020 Team, 2022)
FR-FBn	MF	2008 - 2020	(Warm Winter 2020 Team, 2022)
FR-Fon	DBF	2005 - 2020	(Warm Winter 2020 Team, 2022)
FR-Gri	CRO	2004 - 2020	(Warm Winter 2020 Team, 2022)
FR-Hes	DBF	2014 - 2020	(Warm Winter 2020 Team, 2022)
FR-Lam	ENF	2001 - 2008	(Warm Winter 2020 Team, 2022)
FR-LBr	WET	2017 - 2020	(Berbigier et al., 2016)
FR-LGt	CRO	2005 - 2020	(Warm Winter 2020 Team, 2022)
FR-Pue	EBF	2001 - 2014	(Ourcival et al., 2016)
FR-Tou	GRA	2018 - 2020	(Warm Winter 2020 Team, 2022)
GF-Guy	EBF	2015 - 2015	(Warm Winter 2020 Team, 2022)
GH-Ank	EBF	2011 - 2014	(Valentini et al., 2016a)
GL-NuF	WET	2008 - 2014	(Hansen, 2016)
GL-ZaF	WET	2009 - 2011	(Lund et al., 2016a)
GL-ZaH	GRA	2001 - 2014	(Lund et al., 2016b)
IL-Yat	ENF	2001 - 2020	(Warm Winter 2020 Team, 2022)
IT-CA1	DBF	2011 - 2014	(Sabbatini et al., 2016a)
IT-CA2	CRO	2011 - 2014	(Sabbatini et al., 2016b)
IT-CA3	DBF	2011 - 2014	(Sabbatini et al., 2016c)
IT-Col	DBF	2001 - 2014	(Matteucci, 2016)
IT-Cp2	EBF	2012 - 2020	(Warm Winter 2020 Team, 2022)
IT-Cpz	EBF	2001 - 2008	(Valentini et al., 2016b)
IT-La2	ENF	2001 - 2002	(Cescatti et al., 2016)
IT-Lav	ENF	2003 - 2020	(Warm Winter 2020 Team, 2022)
IT-Lsn	OSH	2016 - 2020	(Warm Winter 2020 Team, 2022)
IT-MBo	GRA	2003 - 2020	(Warm Winter 2020 Team, 2022)
IT-Noe	CSH	2004 - 2014	(Spano et al., 2016)
IT-PT1	DBF	2002 - 2004	(Manca and Goded, 2016)
IT-Ren	ENF	2001 - 2020	(Warm Winter 2020 Team, 2022)
IT-Ro1	DBF	2001 - 2008	(Valentini et al., 2016c)
IT-Ro2	DBF	2002 - 2012	(Papale et al., 2016)
IT-SR2	ENF	2013 - 2020	(Warm Winter 2020 Team, 2022)
IT-SRo	ENF	2001 - 2012	(Gruening et al., 2016)
IT-Tor	GRA	2008 - 2020	(Warm Winter 2020 Team, 2022)
JP-MBF	DBF	2004 - 2005	(Kotani, 2016a)

JP-SMF	MF	2002 - 2006	(Kotani, 2016b)
MY-PSO	EBF	2003 - 2009	(Kosugi and Takanashi, 2016)
NL-Hor	GRA	2004 - 2011	(Dolman et al., 2016a)
NL-Loo	ENF	2001 - 2018	(Drought 2018 Team, 2020)
PA-SPn	DBF	2007 - 2009	(Wolf et al., 2016)
RU-Che	WET	2002 - 2005	(Merbold et al., 2016)
RU-Cok	OSH	2003 - 2013	(Dolman et al., 2016b)
RU-Fy2	ENF	2015 - 2020	(Warm Winter 2020 Team, 2022)
RU-Fyo	ENF	2001 - 2020	(Warm Winter 2020 Team, 2022)
RU-Ha1	GRA	2002 - 2004	(Belelli et al., 2016)
SD-Dem	SAV	2007 - 2009	(Ardö et al., 2016)
SE-Deg	WET	2001 - 2020	(Warm Winter 2020 Team, 2022)
SE-Htm	ENF	2015 - 2020	(Warm Winter 2020 Team, 2022)
SE-Lnn	CRO	2014 - 2018	(Drought 2018 Team, 2020)
SE-Nor	ENF	2014 - 2020	(Warm Winter 2020 Team, 2022)
SE-Ros	ENF	2014 - 2020	(Warm Winter 2020 Team, 2022)
SE-Svb	ENF	2014 - 2020	(Warm Winter 2020 Team, 2022)
SJ-Adv	WET	2013 - 2014	(Christensen, 2016)
SN-Dhr	SAV	2010 - 2013	(Tagesson et al., 2016)
US-ARM	CRO	2004 - 2018	(Biraud et al., 2022)
US-Atq	WET	2003 - 2008	(Zona and Oechel, 2016a)
US-Bar	DBF	2005 - 2017	(Richardson and Hollinger, 2023)
US-Blo	ENF	2001 - 2007	(Goldstein, 2016)
US-Cop	CRO	2011 - 2013	(Bowling, 2016)
US-CRT	GRA	2001 - 2007	(Chen and Chu, 2023)
US-Dk1	GRA	2004 - 2008	(Oishi et al., 2016a)
US-Dk2	DBF	2004 - 2008	(Oishi et al., 2016b)
US-Dk3	ENF	2004 - 2008	(Oishi et al., 2016c)
US-Fmf	WSA	2005 - 2008	(Dore and Kolb, 2023a)
US-FR2	ENF	2005 - 2010	(Litvak, 2016)
US-Fuf	ENF	2005 - 2010	(Dore and Kolb, 2023b)
US-GBT	ENF	2001 - 2003	(Massman, 2016a)
US-GLE	ENF	2005 - 2014	(Massman, 2016b)
US-Goo	GRA	2002 - 2006	(Meyers, 2016)
US-Ha1	DBF	2001 - 2012	(Munger, 2016)
US-Ho1	ENF	2012 - 2018	(Hollinger, 2016)
US-Ivo	WET	2004 - 2007	(Zona and Oechel, 2016b)
US-KFS	GRA	2009 - 2017	(Brunsell, 2022)
US-KS2	CSH	2003 - 2006	(Drake and Hinkle, 2016)
US-Los	WET	2001 - 2014	(Desai, 2016a)
US-Me2	DBF	2001 - 2017	(Law, 2022)
US-Me3	ENF	2003 - 2017	(Law, 2016a)
US-Me5	ENF	2004 - 2009	(Law, 2016b)
US-Me6	ENF	2001 - 2002	(Law, 2016c)
US-MMS	ENF	2010 - 2014	(Novick and Phillips, 2022)
US-Mpj	OSH	2008 - 2017	(Litvak, 2021)
US-Myb	WET	2011 - 2014	(Sturtevant et al., 2016)
US-Ne1	ENF	2001 - 2014	(Suyker, 2016a)
US-Ne2	CRO	2001 - 2013	(Suyker, 2016b)
US-Ne3	CRO	2001 - 2013	(Suyker, 2016c)
US-NR1	CRO	2001 - 2013	(Blanken et al., 2016)
US-Oho	DBF	2004 - 2013	(Chen et al., 2023)

US-PFa	MF	2001 - 2014	(Desai, 2016b)
US-Prr	ENF	2010 - 2016	(Iwahana et al., 2016)
US-Rls	CSH	2014 - 2017	(Flerchinger, 2023)
US-Rms	CSH	2014 - 2017	(Flerchinger, 2022a)
US-Ro1	CRO	2004 - 2016	(Baker et al., 2022)
US-Rws	OSH	2014 - 2017	(Flerchinger, 2022b)
US-Seg	MF	2008 - 2014	(Litvak, 2023a)
US-Ses	WSA	2004 - 2014	(Litvak, 2023b)
US-SRC	GRA	2007 - 2017	(Kurc, 2016)
US-SRM	OSH	2007 - 2017	(Scott, 2016a)
US-Sta	OSH	2005 - 2009	(Ewers and Pendall, 2016)
US-Syv	MF	2001 - 2014	(Desai, 2016c)
US-Ton	WSA	2001 - 2014	(Baldocchi and Ma, 2016)
US-Tw1	WET	2011 - 2017	(Valach et al., 2021)
US-Tw4	WET	2014 - 2017	(Eichelmann et al., 2023)
US-Twt	CRO	2009 - 2014	(Baldocchi, 2016)
US-Uaf	DBF	2007 - 2017	(Ueyama et al., 2018)
US-UMB	DBF	2008 - 2017	(Gough et al., 2023)
US-UMd	ENF	2003 - 2017	(Gough et al., 2022)
US-Var	GRA	2001 - 2014	(Baldocchi et al., 2016)
US-Vcm	ENF	2008 - 2017	(Litvak, 2023c)
US-Vcp	ENF	2007 - 2017	(Litvak, 2023d)
US-WCr	DBF	2001 - 2014	(Desai, 2016d)
US-Whs	WET	2011 - 2013	(Scott, 2016b)
US-Wi3	OSH	2007 - 2014	(Chen, 2016a)
US-Wi4	DBF	2002 - 2004	(Chen, 2016b)
US-Wjs	ENF	2002 - 2005	(Litvak, 2022)
US-WPT	SAV	2007 - 2017	(Chen and Chu, 2016)
ZM-Mon	DBF	2007 - 2009	(Kutsch et al., 2016)

Appendix B: CO_2 sensitivity function of Light Use Efficiency

In the CFE-Hybrid model, the direct CO_2 fertilization effect was prescribed onto machine learning estimated GPP at a reference CO_2 level using a theoretical CO_2 sensitivity function of LUE. The sensitivity function, which describes the fractional change in LUE due to CO_2 relative to the reference period, is described below.

The Light Use Efficiency (LUE) model (Monteith, 1972) of GPP states that,

$$GPP = APAR \times LUE = PAR \times fAPAR \times LUE \tag{A1}$$

where PAR is the photosynthetic active radiation, $fAPAR$ is the fraction of PAR that plant canopy has absorbed, and $APAR$ is the absorbed PAR. Eco-evolutionary theory, specifically the optimal coordination hypothesis, predicts that the electron-transport-limited (light-limited) (A_j) and Rubisco-limited (A_c) rates of photosynthesis converge on the time scale of physiological acclimation, which is in the order of a few weeks (Harrison et al., 2021; Haxeltine and Prentice, 1996; Wang et al., 2017). Thus, at a monthly time scale, we assume that

$$A = A_c = A_j \tag{A2}$$

where A is the gross photosynthetic rate, here equivalent to GPP.

In the following, we derive our sensitivity function based on A_j, which has a smaller response to CO_2 than A_c , thus providing conservative estimates of the direct CO_2 fertilization effect (Walker et al., 2021). According to the Fauquhar, von Caemmerer and Berry (FvCB) model (Farquhar et al., 1980),

$$A_j = \varphi_0 I \frac{c_i - \Gamma^*}{c_i + 2\Gamma^*} \tag{A3}$$

where φ_0 is the intrinsic quantum efficiency of photosynthesis, I is the absorbed PAR ($I = APAR$), c_i is the leaf-internal partial pressure of CO_2, and Γ^* is the photorespiratory compensation point that depends on temperature:

$$\Gamma^* = r_{25} e^{\frac{\Delta H (T - 298.15)}{298.15 RT}} \tag{A4}$$

where $r_{25} = 4.22\ Pa$ is the photorespiratory point at 25 °C, ΔH is the activation energy ($37.83 \cdot 10^3$ J mol^{-1}), T is the air temperature in Kelvin, and R is the molar gas constant (8.314 J mol^{-1} K^{-1}. We denote atmospheric CO_2 concentration as c_a, and χ is the ratio of leaf internal and external CO_2, so

$$c_i = \chi c_a \tag{A5}$$

Combing (A1), (A3), (A5), and assuming (A2), LUE can be written as,

$$LUE = \varphi_0 \frac{c_i - \Gamma^*}{c_i + 2\Gamma^*} = \varphi_0 \frac{\chi c_a - \Gamma^*}{\chi c_a + 2\Gamma^*} \tag{A6}$$

We can therefore show that under constant absorbed light (I or $APAR$), the sensitivity of GPP to CO_2 is proportional to that of LUE,

$$\frac{\partial GPP}{\partial c_a} = \frac{\partial \varphi_0 I \frac{\chi c_a - \Gamma^*}{\chi c_a + 2\Gamma^*}}{\partial c_a} = I \frac{\partial LUE}{\partial c_a} \tag{A7}$$

Thus from (A7), we can express the actual GPP at the time t and a CO_2 level c_a^t as the product of a reference GPP with a CO_2 level c_a^0 and the ratio between actual and reference LUE (A8-9). We denote the actual GPP as time t as $GPP^t_{c_a=c_a^t}$, and the reference GPP at time t as $GPP^t_{c_a=c_a^0}$.

$$\frac{GPP^t_{c_a=c_a^t}}{GPP^t_{c_a=c_a^0}} = \frac{LUE^t_{c_a=c_a^t}}{LUE^t_{c_a=c_a^0}} = \frac{\frac{\chi c_a^t - \Gamma^*}{\chi c_a^t + 2\Gamma^*}}{\frac{\chi c_a^0 - \Gamma^*}{\chi c_a^0 + 2\Gamma^*}} = \frac{\phi^t_{CO2}}{\phi^{t_0}_{CO2}} \tag{A8}$$

$$GPP^t_{c_a=c_a^t} = GPP^t_{c_a=c_a^0} \times \frac{\phi^t_{CO2}}{\phi^{t_0}_{CO2}} \tag{A9}$$

The reference GPP represents the GPP value at time t if the CO_2 were at the level of a reference level, while all other factors, such as PAR, $fAPAR$, temperature, and other environmental controls remain unchanged. Here the CO_2 impacts on LUE depend on atmospheric CO_2 (c_a), χ , and air temperature. We fixed χ to the global long-term average value 0.7 typical to C3 plants (Prentice et al., 2014; Wang et al., 2017). We further tested a dynamic model that quantified χ as a function of air temperature and vapor pressure deficit following an eco-evolutionary theory across global flux sites (Keenan et al., 2023). The estimated χ had a mean and median of 0.7 and a standard deviation of 0.04 (Figure S20~~21~~a). Differences in the direct CO_2 effect between the dynamic and fixed χ approaches were minimal, with an R^2 of 0.99 and a slope of 0.99 from a least squares linear regression line (Figure S20~~21~~b). GPP trends across flux towers were also highly consistent between the two approaches, with a difference less than 0.1 gC m^{-2} $year^{-2}$ (Figure S20~~21~~b, c). Since these results indicated that χ is relatively stable, we used the fixed χ approach to produce the CEDAR-GPP dataset.

In the CFE-Hybrid model, we estimated the reference GPP by fixing the CO_2 at the level of the year 2001 while keeping all other variables dynamic in the CFE-ML model. Then the actual GPP can be estimated following (A9). Fixing CO_2 values to the 2001 level, the start year of eddy covariance data used in model training, essentially removed the effects of CO_2 inferred by the CFE-ML model.

Supplement

The supplement related to this article is available online.

Author contributions

T. K. and Y. K. conceptualized the study. Y. K. performed the formal analysis and generated the final product. Y. K., T. K., M. B., and M. G. contributed to the development and investigation of the research. Y. K., M. G., and X. L. contributed to data curation and processing. Y. K. prepared the manuscript with contributions from all co-authors. T. K. supervised the project.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

We are grateful to Dr. Youngryel Ryu for providing the BESS_Rad dataset and Dr. Martin Jung for sharing the FLUXCOM-RS006 dataset. We also thank Dr. Muyi Li, Dr. Zaichun Zhu, and Dr. Sen Cao for sharing early versions of the PKU GIMMS NDVI4g and LAI4g datasets with us. We extend our gratitude to four reviewers for their constructive feedback for improving this paper.

Financial support

This research was supported by the U.S. Department of Energy Office of Science Early Career Research Program award #DE-SC0021023 and a NASA Award 80NSSC21K1705. YK acknowledges additional support from a NASA award 80NSSC24K1562. TFK acknowledges further support from the LEMONTREE (Land Ecosystem Models based On New Theory, obseRvations and ExperimEnts) project, funded through the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures programme, support from the RUBISCO SFA, which is sponsored by the Regional and Global Model Analysis (RGMA) Program in the Climate and Environmental Sciences Division (CESD) of the Office of Biological and Environmental Research (BER) in the U.S. Department of Energy Office of Science, and NASA awards 80NSSC20K1801 and 80NSSC25K7327. MB acknowledges additional support from the U.S. Department of Agriculture NIFA award #2023-67012-40086.

References

Amiro, B.: FLUXNET2015 CA-Man Manitoba - Northern Old Black Spruce (former BOREAS Northern Study Area), https://doi.org/10.18140/FLX/1440035, 2016a.

Amiro, B.: FLUXNET2015 CA-SF1 Saskatchewan - Western Boreal, forest burned in 1977, https://doi.org/10.18140/FLX/1440046, 2016b.

Amiro, B.: FLUXNET2015 CA-SF3 Saskatchewan - Western Boreal, forest burned in 1998, https://doi.org/10.18140/FLX/1440048, 2016c.

Amiro, B.: AmeriFlux FLUXNET-1F CA-SF2 Saskatchewan - Western Boreal, forest burned in 1989, https://doi.org/10.17190/AMF/2006961, 2023.

Ammann, C.: FLUXNET2015 CH-Oe1 Oensingen grassland, https://doi.org/10.18140/FLX/1440135, 2016.

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, 1–34, https://doi.org/10.1002/2015RG000483, 2015.

Apley, D. W. and Zhu, J.: Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models, Journal of the Royal Statistical Society Series B: Statistical Methodology, 82, 1059–1086, https://doi.org/10.1111/rssb.12377, 2020.

Arain, M. A.: AmeriFlux AmeriFlux CA-TP4 Ontario - Turkey Point 1939 Plantation White Pine, https://doi.org/10.17190/AMF/1246012, 2016a.

Arain, M. A.: FLUXNET2015 CA-TP1 Ontario - Turkey Point 2002 Plantation White Pine, https://doi.org/10.18140/FLX/1440050, 2016b.

Arain, M. A.: FLUXNET2015 CA-TP2 Ontario - Turkey Point 1989 Plantation White Pine, https://doi.org/10.18140/FLX/1440051, 2016c.

Arain, M. A.: FLUXNET2015 CA-TP3 Ontario - Turkey Point 1974 Plantation White Pine, https://doi.org/10.18140/FLX/1440052, 2016d.

Arain, M. A.: FLUXNET2015 CA-TPD Ontario - Turkey Point Mature Deciduous, https://doi.org/10.18140/FLX/1440112, 2016e.

Ardö, J., El Tahir, B. A., and ElKhidir, H. A. M.: FLUXNET2015 SD-Dem Demokeya, https://doi.org/10.18140/FLX/1440186, 2016.

Arndt, S., Hinko-Najera, N., Griebel, A., Beringer, J., and Livesley, S. J.: FLUXNET2015 AU-Wom Wombat, https://doi.org/10.18140/FLX/1440207, 2016.

Aurela, M., Lohila, A., Tuovinen, J.-P., Hatakka, J., Rainne, J., Mäkelä, T., and Lauria, T.: FLUXNET2015 FI-Lom Lompolojankka, https://doi.org/10.18140/FLX/1440228, 2016a.

Aurela, M., Tuovinen, J.-P., Hatakka, J., Lohila, A., Mäkelä, T., Rainne, J., and Lauria, T.: FLUXNET2015 FI-Sod Sodankyla, https://doi.org/10.18140/FLX/1440160, 2016b.

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

Baker, J., Griffis, T., and Griffis, T.: AmeriFlux FLUXNET-1F US-Ro1 Rosemount- G21, https://doi.org/10.17190/AMF/1881588, 2022.

Baldocchi, D.: FLUXNET2015 US-Twt Twitchell Island, https://doi.org/10.18140/FLX/1440106, 2016.

Baldocchi, D. and Ma, S.: FLUXNET2015 US-Ton Tonzi Ranch, https://doi.org/10.18140/FLX/1440092, 2016.

Baldocchi, D., Ma, S., and Xu, L.: FLUXNET2015 US-Var Vaira Ranch- Ione, https://doi.org/10.18140/FLX/1440094, 2016.

Baldocchi, D., Chu, H., and Reichstein, M.: Inter-annual variability of net and gross ecosystem carbon fluxes: A review, Agricultural and Forest Meteorology, 249, 520–533, https://doi.org/10.1016/j.agrformet.2017.05.015, 2018.

Baldocchi, D. D.: How eddy covariance flux measurements have contributed to our understanding of Global Change Biology, Global Change Biology, 26, 242–260, https://doi.org/10.1111/gcb.14807, 2020.

Baniecki, H., Kretowicz, W., Piątyszek, P., Wiśniewski, J., and Biecek, P.: dalex: Responsible Machine Learning with Interactive Explainability and Fairness in Python, Journal of Machine Learning Research, 22, 1–7, 2021.

Barr, A. and Black, A. T.: AmeriFlux AmeriFlux CA-SJ2 Saskatchewan - Western Boreal, Jack Pine forest harvested in 2002, https://doi.org/10.17190/AMF/1436321, 2018.

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., and Wood, E. F.: Present and future köppen-geiger climate classification maps at 1-km resolution, Scientific Data, 5, 1–12, https://doi.org/10.1038/sdata.2018.214, 2018.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M. A., Baldocchi, D., Bonan, G. B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K. W., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F. I., and Papale, D.: Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate, Science, 329, 834–838, https://doi.org/10.1126/science.1184984, 2010.

Belelli, L., Papale, D., and Valentini, R.: FLUXNET2015 RU-Ha1 Hakasia steppe, https://doi.org/10.18140/FLX/1440184, 2016.

Berbigier, P., Loustau, D., Bonnefond, J. M., Bosc, A., and Trichet, P.: FLUXNET2015 FR-LBr Le Bray, https://doi.org/10.18140/FLX/1440163, 2016.

Berdugo, M., Gaitán, J. J., Delgado-Baquerizo, M., Crowther, T. W., and Dakos, V.: Prevalence and drivers of abrupt vegetation shifts in global drylands, Proceedings of the National Academy of Sciences, 119, e2123393119, https://doi.org/10.1073/pnas.2123393119, 2022.

Beringer, J. and Hutley, L.: FLUXNET2015 AU-Ade Adelaide River, https://doi.org/10.18140/FLX/1440193, 2016a.

Beringer, J. and Hutley, L.: FLUXNET2015 AU-DaP Daly River Savanna, https://doi.org/10.18140/FLX/1440123, 2016b.

Beringer, J. and Hutley, L.: FLUXNET2015 AU-Dry Dry River, https://doi.org/10.18140/FLX/1440197, 2016c.

Beringer, J. and Hutley, L.: FLUXNET2015 AU-Fog Fogg Dam, https://doi.org/10.18140/FLX/1440124, 2016d.

Beringer, J. and Hutley, L.: FLUXNET2015 AU-How Howard Springs, https://doi.org/10.18140/FLX/1440125, 2016e.

Beringer, J. and Hutley, L.: FLUXNET2015 AU-RDF Red Dirt Melon Farm, Northern Territory, https://doi.org/10.18140/FLX/1440201, 2016f.

Beringer, J. and Hutley, P. L.: FLUXNET2015 AU-DaS Daly River Cleared, https://doi.org/10.18140/FLX/1440122, 2016g.

Beringer, J. and Walker, J.: FLUXNET2015 AU-Ync Jaxa, https://doi.org/10.18140/FLX/1440208, 2016.

Beringer, J., Cunningham, S., Baker, P., Cavagnaro, T., MacNally, R., Thompson, R., and McHugh, I.: FLUXNET2015 AU-Rig Riggs Creek, https://doi.org/10.18140/FLX/1440202, 2016a.

Beringer, J., Hutley, L., McGuire, D., U, P., and McHugh, I.: FLUXNET2015 AU-Wac Wallaby Creek, https://doi.org/10.18140/FLX/1440127, 2016b.

Beringer, J., Cunningham, S., Baker, P., Cavagnaro, T., MacNally, R., Thompson, R., and McHugh, I.: FLUXNET2015 AU-Whr Whroo, https://doi.org/10.18140/FLX/1440206, 2016c.

Bernhofer, C., Grünwald, T., Moderow, U., Hehn, M., Eichelmann, U., Prasse, H., and Postel, U.: FLUXNET2015 DE-Spw Spreewald, https://doi.org/10.18140/FLX/1440220, 2016.

Besnard, S., Carvalhais, N., Altaf Arain, M., Black, A., Brede, B., Buchmann, N., Chen, J., Clevers, J. G. P. W., Dutrieux, L. P., Gans, F., Herold, M., Jung, M., Kosugi, Y., Knohl, A., Law, B. E., Paul-Limoges, E., Lohila, A., Merbold, L., Roupsard, O., Valentini, R., Wolf, S., Zhang, X., and Reichstein, M.: Memory effects of climate and vegetation affecting net ecosystem CO2 fluxes in global forests, PLoS ONE, 14, 1–22, https://doi.org/10.1371/journal.pone.0211510, 2019.

Biraud, S., Fischer, M., Chan, S., and Torn, M.: AmeriFlux FLUXNET-1F US-ARM ARM Southern Great Plains site- Lamont, https://doi.org/10.17190/AMF/1854366, 2022.

Black, T. A.: FLUXNET2015 CA-Oas Saskatchewan - Western Boreal, Mature Aspen, https://doi.org/10.18140/FLX/1440043, 2016a.

Black, T. A.: FLUXNET2015 CA-Obs Saskatchewan - Western Boreal, Mature Black Spruce, https://doi.org/10.18140/FLX/1440044, 2016b.

Black, T. A.: AmeriFlux AmeriFlux CA-Ca3 British Columbia - Pole sapling Douglas-fir stand, https://doi.org/10.17190/AMF/1480302, 2018.

Black, T. A.: AmeriFlux FLUXNET-1F CA-Ca1 British Columbia - 1949 Douglas-fir stand, https://doi.org/10.17190/AMF/2007163, 2023a.

Black, T. A.: AmeriFlux FLUXNET-1F CA-Ca2 British Columbia - Clearcut Douglas-fir stand (harvested winter 1999/2000), https://doi.org/10.17190/AMF/2007164, 2023b.

Blanken, P. D., Monson, R. K., Burns, S. P., Bowling, D. R., and Turnipseed, A. A.: FLUXNET2015 US-NR1 Niwot Ridge Forest (LTER NWT1), https://doi.org/10.18140/FLX/1440087, 2016.

Bloomfield, K. J., Stocker, B. D., Keenan, T. F., and Prentice, I. C.: Environmental controls on the light use efficiency of terrestrial gross primary production, Global Change Biology, 29, 1037–1053, https://doi.org/10.1111/gcb.16511, 2023.

Bowling, D.: FLUXNET2015 US-Cop Corral Pocket, https://doi.org/10.18140/FLX/1440100, 2016.

Brunsell, N.: AmeriFlux FLUXNET-1F US-KFS Kansas Field Station, https://doi.org/10.17190/AMF/1881585, 2022.

Campbell, J. E., Berry, J. A., Seibt, U., Smith, S. J., Montzka, S. A., Launois, T., Belviso, S., Bopp, L., and Laine, M.: Large historical growth in global terrestrial gross primary production, Nature, 544, 84–87, https://doi.org/10.1038/nature22030, 2017.

Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walther, S., Duveiller, G., Cescatti, A., Mahecha, M. D., Muñoz-Marí, J., García-Haro, F. J., Guanter, L., Jung, M., Gamon, J. A., Reichstein, M., and Running, S. W.: A unified vegetation index for quantifying the terrestrial biosphere, Science Advances, 7, eabc7447, https://doi.org/10.1126/sciadv.abc7447, 2021.

Cao, S., Li, M., Zhu, Z., Zha, J., Zhao, W., Duanmu, Z., Chen, J., Zheng, Y., and Chen, Y.: Spatiotemporally consistent global dataset of the GIMMS Leaf Area Index (GIMMS LAI4g) from 1982 to 2020, Earth System Science Data Discussions, 1–31, https://doi.org/10.5194/essd-2023-68, 2023.

Cescatti, A., Marcolla, B., Zorer, R., and Gianelle, D.: FLUXNET2015 IT-La2 Lavarone2, https://doi.org/10.18140/FLX/1440235, 2016.

Chen, C., Park, T., Wang, X., Piao, S., Xu, B., Chaturvedi, R. K., Fuchs, R., Brovkin, V., Ciais, P., Fensholt, R., Tømmervik, H., Bala, G., Zhu, Z., Nemani, R. R., and Myneni, R. B.: China and India lead in greening of the world through land-use management, Nature Sustainability, 2, 122–129, https://doi.org/10.1038/s41893-019-0220-7, 2019.

Chen, C., Riley, W. J., Prentice, I. C., and Keenan, T. F.: CO2 fertilization of terrestrial photosynthesis inferred from site to global scales, Proceedings of the National Academy of Sciences, 119, 1–8, https://doi.org/10.1073/pnas.2115627119/, 2022.

Chen, J.: FLUXNET2015 US-Wi3 Mature hardwood (MHW), https://doi.org/10.18140/FLX/1440057, 2016a.

Chen, J.: FLUXNET2015 US-Wi4 Mature red pine (MRP), https://doi.org/10.18140/FLX/1440058, 2016b.

Chen, J. and Chu, H.: FLUXNET2015 US-WPT Winous Point North Marsh, https://doi.org/10.18140/FLX/1440116, 2016.

Chen, J. and Chu, H.: AmeriFlux FLUXNET-1F US-CRT Curtice Walter-Berger cropland, https://doi.org/10.17190/AMF/2006974, 2023.

Chen, J., Chu, H., and Noormets, A.: AmeriFlux FLUXNET-1F US-Oho Oak Openings, https://doi.org/10.17190/AMF/2229385, 2023.

Chen, S.: FLUXNET2015 CN-Du2 Duolun_grassland (D01), https://doi.org/10.18140/FLX/1440140, 2016c.

Chen, T. and Guestrin, C.: XGBoost: a scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, 785–794, https://doi.org/10.1145/2939672.2939785, 2016.

Christensen, T.: FLUXNET2015 SJ-Adv Adventdalen, https://doi.org/10.18140/FLX/1440241, 2016.

Chu, H., Luo, X., Ouyang, Z., Chan, W. S., Dengel, S., Biraud, S. C., Torn, M. S., Metzger, S., Kumar, J., Arain, M. A., Arkebauer, T. J., Baldocchi, D., Bernacchi, C., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Bracho, R., Brown, S., Brunsell, N. A., Chen, J., Chen, X., Clark, K., Desai, A. R., Duman, T., Durden, D., Fares, S., Forbrich, I., Gamon, J. A., Gough, C. M., Griffis, T., Helbig, M., Hollinger, D., Humphreys, E., Ikawa, H., Iwata, H., Ju, Y., Knowles, J. F., Knox, S. H., Kobayashi, H., Kolb, T., Law, B., Lee, X., Litvak, M., Liu, H., Munger, J. W., Noormets, A., Novick, K., Oberbauer, S. F., Oechel, W., Oikawa, P., Papuga, S. A., Pendall, E., Prajapati, P., Prueger, J., Quinton, W. L., Richardson, A. D., Russell, E. S., Scott, R. L., Starr, G., Staebler, R., Stoy, P. C., Stuart-Haëntjens, E., Sonnentag, O., Sullivan, R. C., Suyker, A., Ueyama, M., Vargas, R., Wood, J. D., and Zona, D.: Representativeness of Eddy-Covariance flux footprints for areas surrounding AmeriFlux sites, Agricultural and Forest Meteorology, 301–302, https://doi.org/10.1016/j.agrformet.2021.108350, 2021.

Cleverly, J., Eamus, D., and Isaac, P.: FLUXNET2015 AU-ASM Alice Springs, https://doi.org/10.18140/FLX/1440194, 2016.

Dannenberg, M. P., Barnes, M. L., Smith, W. K., Johnston, M. R., Meerdink, S. K., Wang, X., Scott, R. L., and Biederman, J. A.: Upscaling dryland carbon and water fluxes with artificial neural networks of optical, thermal, and microwave satellite remote sensing, Biogeosciences, 20, 383–404, https://doi.org/10.5194/bg-20-383-2023, 2023.

De Kauwe, M. G., Keenan, T. F., Medlyn, B. E., Prentice, I. C., and Terrer, C.: Satellite based estimates underestimate the effect of CO2 fertilization on net primary productivity, Nature Climate Change, 6, 892–893, https://doi.org/10.1038/nclimate3105, 2016.

Delwiche, K. B., Nelson, J., Kowalska, N., Moore, C. E., Shirkey, G., Tarin, T., Cleverly, J. R., and Keenan, T. F.: Charting the Future of the FLUXNET Network, Bulletin of the American Meteorological Society, 105, E466–E473, https://doi.org/10.1175/BAMS-D-23-0316.1, 2024.

Desai, A.: FLUXNET2015 US-Los Lost Creek, https://doi.org/10.18140/FLX/1440076, 2016a.

Desai, A.: FLUXNET2015 US-PFa Park Falls/WLEF, https://doi.org/10.18140/FLX/1440089, 2016b.

Desai, A.: FLUXNET2015 US-Syv Sylvania Wilderness Area, https://doi.org/10.18140/FLX/1440091, 2016c.

Desai, A.: FLUXNET2015 US-WCr Willow Creek, https://doi.org/10.18140/FLX/1440095, 2016d.

Dolman, H., Hendriks, D., Parmentier, F.-J., Marchesini, L. B., Dean, J., and van Huissteden, K.: FLUXNET2015 NL-Hor Horstermeer, https://doi.org/10.18140/FLX/1440177, 2016a.

Dolman, H., van der Molen, M., Parmentier, F.-J., Marchesini, L. B., Dean, J., van Huissteden, K., and Maximov, T.: FLUXNET2015 RU-Cok Chokurdakh, https://doi.org/10.18140/FLX/1440182, 2016b.

Dong, G.: FLUXNET2015 CN-Cng Changling, https://doi.org/10.18140/FLX/1440209, 2016.

Dong, Y., Li, J., Jiao, Z., Liu, Q., Zhao, J., Xu, B., Zhang, H., Zhang, Z., Liu, C., Knyazikhin, Y., and Myneni, R. B.: A Method for Retrieving Coarse-Resolution Leaf Area Index for Mixed Biomes Using a Mixed-Pixel Correction Factor, IEEE Transactions on Geoscience and Remote Sensing, 61, 1–17, https://doi.org/10.1109/TGRS.2023.3235949, 2023.

Dore, S. and Kolb, T.: AmeriFlux FLUXNET-1F US-Fmf Flagstaff - Managed Forest, https://doi.org/10.17190/AMF/2007173, 2023a.

Dore, S. and Kolb, T.: AmeriFlux FLUXNET-1F US-Fuf Flagstaff - Unmanaged Forest, https://doi.org/10.17190/AMF/2007174, 2023b.

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P. D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y. Y., Miralles, D., Mistelbauer, T., Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimer, C., van der Schalie, R., Seneviratne, S. I., Smolander, T., and Lecomte, P.: ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions, Remote Sensing of Environment, 203, 185–215, https://doi.org/10.1016/j.rse.2017.07.001, 2017.

Dorigo, W. A., Gruber, A., De Jeu, R. A. M., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R. M., and Kidd, R.: Evaluation of the ESA CCI soil moisture product using ground-based observations, Remote Sensing of Environment, 162, https://doi.org/10.1016/j.rse.2014.07.023, 2015.

Drake, B. and Hinkle, R.: FLUXNET2015 US-KS2 Kennedy Space Center (scrub oak), https://doi.org/10.18140/FLX/1440075, 2016.

Drought 2018 Team: Drought-2018 ecosystem eddy covariance flux product for 52 stations in FLUXNET-Archive format, https://doi.org/10.18160/YVR0-4898, 2020.

Ehlers, I., Augusti, A., Betson, T. R., Nilsson, M. B., Marshall, J. D., and Schleucher, J.: Detecting long-term metabolic shifts using isotopomers: CO2-driven suppression of photorespiration in C3 plants over the 20th century, Proceedings of the National Academy of Sciences, 112, 15585–15590, https://doi.org/10.1073/pnas.1504493112, 2015.

Eichelmann, E., Shortt, R., Knox, S., Sanchez, C. R., Valach, A., Sturtevant, C., Szutu, D., Verfaillie, J., and Baldocchi, D.: AmeriFlux FLUXNET-1F US-Tw4 Twitchell East End Wetland, https://doi.org/10.17190/AMF/2204881, 2023.

Ewers, B. and Pendall, E.: FLUXNET2015 US-Sta Saratoga, https://doi.org/10.18140/FLX/1440115, 2016.

Fang, H., Baret, F., Plummer, S., and Schaepman-Strub, G.: An overview of global leaf area index (LAI): Methods, products, validation, and applications, Reviews of Geophysics, 2018RG000608, https://doi.org/10.1029/2018RG000608, 2019.

Farquhar, G. D., von Caemmerer, S., and Berry, J. A.: A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species, Planta, 149, 78–90, https://doi.org/10.1007/BF00386231, 1980.

Fernández-Martínez, M., Vicca, S., Janssens, I. A., Ciais, P., Obersteiner, M., Bartrons, M., Sardans, J., Verger, A., Canadell, J. G., Chevallier, F., Wang, X., Bernhofer, C., Curtis, P. S., Gianelle, D., Grünwald, T., Heinesch, B., Ibrom, A., Knohl, A., Laurila, T., Law, B. E., Limousin, J. M., Longdoz, B., Loustau, D., Mammarella, I., Matteucci, G., Monson, R. K., Montagnani, L., Moors, E. J., Munger, J. W., Papale, D., Piao, S. L., and Peñuelas, J.: Atmospheric deposition, CO2, and change in the land carbon sink, Sci Rep, 7, 9632, https://doi.org/10.1038/s41598-017-08755-8, 2017.

Flanagan, L. B.: AmeriFlux AmeriFlux CA-WP1 Alberta - Western Peatland - LaBiche River,Black Spruce/Larch Fen, https://doi.org/10.17190/AMF/1436323, 2018a.

Flanagan, L. B.: AmeriFlux AmeriFlux CA-WP2 Alberta - Western Peatland - Poor Fen (Sphagnum moss), https://doi.org/10.17190/AMF/1436324, 2018b.

Flanagan, L. B.: AmeriFlux AmeriFlux CA-WP3 Alberta - Western Peatland - Rich Fen (Carex), https://doi.org/10.17190/AMF/1436325, 2018c.

Flerchinger, G.: AmeriFlux FLUXNET-1F US-Rms RCEW Mountain Big Sagebrush, https://doi.org/10.17190/AMF/1881587, 2022a.

Flerchinger, G.: AmeriFlux FLUXNET-1F US-Rws Reynolds Creek Wyoming big sagebrush, https://doi.org/10.17190/AMF/1881592, 2022b.

Flerchinger, G.: AmeriFlux FLUXNET-1F US-Rls RCEW Low Sagebrush, https://doi.org/10.17190/AMF/2229387, 2023.

Friedl, M. and Sulla-Menashe, D.: MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006 [Data set], NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/MODIS/MCD12Q1.006, 2019.

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., and Knutti, R.: Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks, Journal of Climate, 27, 511–526, https://doi.org/10.1175/JCLI-D-12-00579.1, 2014.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., Le Quéré, C., Luijkx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Alkama, R., Arneth, A., Arora, V. K., Bates, N. R., Becker, M., Bellouin, N., Bittig, H. C., Bopp, L., Chevallier, F., Chini, L. P., Cronin, M., Evans, W., Falk, S., Feely, R. A., Gasser, T., Gehlen, M., Gkritzalis, T., Gloege, L., Grassi, G., Gruber, N., Gürses, Ö., Harris, I., Hefner, M., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Jain, A. K., Jersild, A., Kadono, K., Kato, E., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Landschützer, P., Lefèvre, N., Lindsay, K., Liu, J., Liu, Z., Marland, G., Mayot, N., McGrath, M. J., Metzl, N., Monacci, N. M., Munro, D. R., Nakaoka, S.-I., Niwa, Y., O'Brien, K., Ono, T., Palmer, P. I., Pan, N., Pierrot, D., Pocock, K., Poulter, B., Resplandy, L., Robertson, E., Rödenbeck, C., Rodriguez, C., Rosan, T. M., Schwinger, J., Séférian, R., Shutler, J. D., Skjelvan, I., Steinhoff, T., Sun, Q., Sutton, A. J., Sweeney, C., Takao, S., Tanhua, T., Tans, P. P., Tian, X., Tian, H., Tilbrook, B., Tsujino, H., Tubiello, F., van der Werf, G. R., Walker, A. P., Wanninkhof, R., Whitehead, C., Willstrand Wranne, A., et al.: Global Carbon Budget 2022, Earth System Science Data, 14, 4811–4900, 2023.

Gaber, M., Kang, Y., Schurgers, G., and Keenan, T.: Using automated machine learning for the upscaling of gross primary productivity, Biogeosciences, 21, 2447–2472, https://doi.org/10.5194/bg-21-2447-2024, 2024.

Gampe, D., Zscheischler, J., Reichstein, M., Sullivan, M. O., Smith, W. K., Sitch, S., and Buermann, W.: Increasing impact of warm droughts on northern ecosystem productivity over recent decades, Nature Climate Change, https://doi.org/10.1038/s41558-021-01112-8, 2021.

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

Garcia, A., Di Bella, C., Houspanossian, J., Magliano, P., Jobbágy, E., Posse, G., Fernández, R., and Nosetto, M.: FLUXNET2015 AR-SLu San Luis, 2016.

Gitelson, A. A.: Remote estimation of leaf area index and green leaf biomass in maize canopies, Geophysical Research Letters, 30, 1248, https://doi.org/10.1029/2002GL016450, 2003.

Goldstein, A.: FLUXNET2015 US-Blo Blodgett Forest, https://doi.org/10.18140/FLX/1440068, 2016.

Gough, C., Bohrer, G., and Curtis, P.: AmeriFlux FLUXNET-1F US-UMd UMBS Disturbance, https://doi.org/10.17190/AMF/1881597, 2022.

Gough, C., Bohrer, G., and Curtis, P.: AmeriFlux FLUXNET-1F US-UMB Univ. of Mich. Biological Station, https://doi.org/10.17190/AMF/2204882, 2023.

Goulden, M.: FLUXNET2015 BR-Sa3 Santarem-Km83-Logged Forest, https://doi.org/10.18140/FLX/1440033, 2016a.

Goulden, M.: FLUXNET2015 CA-NS4 UCI-1964 burn site wet, https://doi.org/10.18140/FLX/1440039, 2016b.

Goulden, M.: FLUXNET2015 CA-NS7 UCI-1998 burn site, https://doi.org/10.18140/FLX/1440042, 2016c.

Goulden, M.: AmeriFlux FLUXNET-1F CA-NS1 UCI-1850 burn site, https://doi.org/10.17190/AMF/1902824, 2022a.

Goulden, M.: AmeriFlux FLUXNET-1F CA-NS2 UCI-1930 burn site, https://doi.org/10.17190/AMF/1902825, 2022b.

Goulden, M.: AmeriFlux FLUXNET-1F CA-NS3 UCI-1964 burn site, https://doi.org/10.17190/AMF/1902826, 2022c.

Goulden, M.: AmeriFlux FLUXNET-1F CA-NS5 UCI-1981 burn site, https://doi.org/10.17190/AMF/1902828, 2022d.

Goulden, M.: AmeriFlux FLUXNET-1F CA-NS6 UCI-1989 burn site, https://doi.org/10.17190/AMF/1902829, 2022e.

Green, J. K., Ballantyne, A., Abramoff, R., Gentine, P., Makowski, D., and Ciais, P.: Surface temperatures reveal the patterns of vegetation water stress and their environmental drivers across the tropical Americas, Global Change Biology, 28, 2940–2955, https://doi.org/10.1111/gcb.16139, 2022.

Gruber, A., Scanlon, T., Van Der Schalie, R., Wagner, W., and Dorigo, W.: Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology, Earth System Science Data, 11, 717–739, https://doi.org/10.5194/essd-11-717-2019, 2019.

Gruening, C., Goded, I., Cescatti, A., Manca, G., and Seufert, G.: FLUXNET2015 IT-SRo San Rossore, https://doi.org/10.18140/FLX/1440176, 2016.

Hansen, B. U.: FLUXNET2015 GL-NuF Nuuk Fen, https://doi.org/10.18140/FLX/1440222, 2016.

Hao, D., Bisht, G., Huang, M., Ma, P.-L., Tesfa, T., Lee, W.-L., Gu, Y., and Leung, L. R.: Impacts of Sub-Grid Topographic Representations on Surface Energy Balance and Boundary Conditions in the E3SM Land Model: A Case Study in Sierra Nevada, Journal of Advances in Modeling Earth Systems, 14, e2021MS002862, https://doi.org/10.1029/2021MS002862, 2022.

Harrison, S. P., Cramer, W., Franklin, O., Prentice, I. C., Wang, H., Brännström, Å., de Boer, H., Dieckmann, U., Joshi, J., Keenan, T. F., Lavergne, A., Manzoni, S., Mengoli, G., Morfopoulos, C., Peñuelas, J., Pietsch, S., Rebel, K. T., Ryu, Y., Smith, N. G., Stocker, B. D., and Wright, I. J.: Eco-evolutionary optimality as a means to improve vegetation and land-surface models, New Phytologist, https://doi.org/10.1111/nph.17558, 2021.

Haverd, V., Smith, B., Canadell, J. G., Cuntz, M., Mikaloff-Fletcher, S., Farquhar, G., Woodgate, W., Briggs, P. R., and Trudinger, C. M.: Higher than expected CO_2 fertilization inferred from leaf to global observations, Global Change Biology, 26, 2390–2402, https://doi.org/10.1111/gcb.14950, 2020.

Haxeltine, A. and Prentice, I. C.: A General Model for the Light-Use Efficiency of Primary Production, Functional Ecology, 10, 551–561, https://doi.org/10.2307/2390165, 1996.

Hollinger, D.: AmeriFlux AmeriFlux US-Ho1 Howland Forest (main tower), https://doi.org/10.17190/AMF/1246061, 2016.

van der Horst, S. V. J., Pitman, A. J., De Kauwe, M. G., Ukkola, A., Abramowitz, G., and Isaac, P.: How representative are FLUXNET measurements of surface fluxes during temperature extremes?, Biogeosciences, 16, 1829–1844, https://doi.org/10.5194/bg-16-1829-2019, 2019.

Iwahana, G., Kobayashi, H., Ikawa, H., and Suzuki, R.: AmeriFlux AmeriFlux US-Prr Poker Flat Research Range Black Spruce Forest, https://doi.org/10.17190/AMF/1246153, 2016.

Jiang, C., Ryu, Y., Fang, H., Myneni, R., Claverie, M., and Zhu, Z.: Inconsistencies of interannual variability and trends in long-term satellite leaf area index products, Global Change Biology, 23, 4133–4146, https://doi.org/10.1111/gcb.13787, 2017.

Joiner, J. and Yoshida, Y.: Satellite-based reflectances capture large fraction of variability in global gross primary production (GPP) at weekly time scales, Agricultural and Forest Meteorology, 291, 108092, https://doi.org/10.1016/j.agrformet.2020.108092, 2020.

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F., and Williams, C.: Global patterns of land-atmosphere fluxes of carbon dioxide , latent heat , and sensible heat derived from eddy covariance , satellite , and meteorological observations, Journal of Geophysical Research: Biogeosciences, 116, 1–16, https://doi.org/10.1029/2010JG001566, 2011.

Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., Gans, F., Ichii, K., Jain, A. K., Kato, E., Papale, D., Poulter, B., Raduly, B., Rödenbeck, C., Tramontana, G., Viovy, N., Wang, Y. P., Weber, U., Zaehle, S., and Zeng, N.: Compensatory water effects link yearly global land CO_2 sink changes to temperature, Nature, 541, 516–520, https://doi.org/10.1038/nature20780, 2017.

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

Kang, Y., Ozdogan, M., Zhu, X., Ye, Z., Hain, C., and Anderson, M.: Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest, Environmental Research Letters, 15, https://doi.org/10.1088/1748-9326/ab7df9, 2020.

Kang, Y., Bassiouni, M., Gaber, M., Lu, X., and Keenan, T.: CEDAR-GPP: A Spatiotemporally Upscaled Dataset of Gross Primary Productivity Incorporating CO2 Fertilization (v1.0), https://doi.org/10.5281/zenodo.8212706, 2024.

Keeling, R. F., Graven, H. D., Welp, L. R., Resplandy, L., Bi, J., Piper, S. C., Sun, Y., Bollenbacher, A., and Meijer, H. A. J.: Atmospheric evidence for a global secular increase in carbon isotopic discrimination of land photosynthesis, Proceedings of the National Academy of Sciences of the United States of America, 114, 10361–10366, https://doi.org/10.1073/pnas.1619240114, 2017.

Keenan, T. F., Hollinger, D. Y., Bohrer, G., Dragoni, D., Munger, J. W., Schmid, H. P., and Richardson, A. D.: Increase in forest water-use efficiency as atmospheric carbon dioxide concentrations rise, Nature, 499, 324–327, https://doi.org/10.1038/nature12291, 2013.

Keenan, T. F., Prentice, I. C., Canadell, J. G., Williams, C. A., Wang, H., Raupach, M., and Collatz, G. J.: Recent pause in the growth rate of atmospheric CO2 due to enhanced terrestrial carbon uptake, Nature Communications, 7, 1–9, https://doi.org/10.1038/ncomms13428, 2016.

Keenan, T. F., Migliavacca, M., Papale, D., Baldocchi, D., Reichstein, M., Torn, M., and Wutzler, T.: Widespread inhibition of daytime ecosystem respiration, Nature Ecology and Evolution, 3, 407–415, https://doi.org/10.1038/s41559-019-0809-2, 2019.

Keenan, T. F., Luo, X., Stocker, B. D., De Kauwe, M. G., Medlyn, B. E., Prentice, I. C., Smith, N. G., Terrer, C., Wang, H., Zhang, Y., and Zhou, S.: A constraint on historic growth in global photosynthesis due to rising CO2, Nat. Clim. Chang., 1–6, https://doi.org/10.1038/s41558-023-01867-2, 2023.

Klatt, J., Schmid, H. P., Mauder, M., and Steinbrecher, R.: FLUXNET2015 DE-SfN Schechenfilz Nord, https://doi.org/10.18140/FLX/1440219, 2016.

Knohl, A., Tiedemann, F., Kolle, O., Schulze, E.-D., Anthoni, P., Kutsch, W., Herbst, M., and Siebicke, L.: FLUXNET2015 DE-Lnf Leinefelde, https://doi.org/10.18140/FLX/1440150, 2016.

Kosugi, Y. and Takanashi, S.: FLUXNET2015 MY-PSO Pasoh Forest Reserve (PSO), https://doi.org/10.18140/FLX/1440240, 2016.

Kotani, A.: FLUXNET2015 JP-MBF Moshiri Birch Forest Site, https://doi.org/10.18140/FLX/1440238, 2016a.

Kotani, A.: FLUXNET2015 JP-SMF Seto Mixed Forest Site, https://doi.org/10.18140/FLX/1440239, 2016b.

Kraft, B., Jung, M., Körner, M., Koirala, S., and Reichstein, M.: Towards hybrid modeling of the global hydrological cycle, Hydrology and Earth System Sciences, 26, 1579–1614, https://doi.org/10.5194/hess-26-1579-2022, 2022.

Kurc, S.: FLUXNET2015 US-SRC Santa Rita Creosote, https://doi.org/10.18140/FLX/1440098, 2016.

Kutsch, W. L., Merbold, L., and Kolle, O.: FLUXNET2015 ZM-Mon Mongu, https://doi.org/10.18140/FLX/1440189, 2016.

Law, B.: FLUXNET2015 US-Me3 Metolius-second young aged pine, https://doi.org/10.18140/FLX/1440080, 2016a.

Law, B.: FLUXNET2015 US-Me5 Metolius-first young aged pine, https://doi.org/10.18140/FLX/1440082, 2016b.

Law, B.: FLUXNET2015 US-Me6 Metolius Young Pine Burn, https://doi.org/10.18140/FLX/1440099, 2016c.

Law, B.: AmeriFlux FLUXNET-1F US-Me2 Metolius mature ponderosa pine, https://doi.org/10.17190/AMF/1854368, 2022.

Leng, J., Chen, J. M., Li, W., Luo, X., Xu, M., Liu, J., Wang, R., Rogers, C., Li, B., and Yan, Y.: Global datasets of hourly carbon and water fluxes simulated using a satellite-based process model with dynamic parameterizations, Earth System Science Data, 16, 1283–1300, https://doi.org/10.5194/essd-16-1283-2024, 2024.

Li, B., Ryu, Y., Jiang, C., Dechant, B., Liu, J., Yan, Y., and Li, X.: BESSv2.0: A satellite-based and coupled-process model for quantifying long-term global land–atmosphere fluxes, Remote Sensing of Environment, 295, 113696, https://doi.org/10.1016/j.rse.2023.113696, 2023a.

Li, M., Cao, S., and Zhu, Z.: Spatiotemporally consistent global dataset of the GIMMS Normalized Difference Vegetation Index (PKU GIMMS NDVI) from 1982 to 2020, Earth System Science Data Discussions, 1–31, https://doi.org/10.5194/essd-2023-1, 2023b.

Li, Y.: FLUXNET2015 CN-Ha2 Haibei Shrubland, https://doi.org/10.18140/FLX/1440211, 2016.

Lindauer, M., Steinbrecher, R., Wolpert, B., Mauder, M., and Schmid, H. P.: FLUXNET2015 DE-Lkb Lackenberg, https://doi.org/10.18140/FLX/1440214, 2016.

Litvak, M.: AmeriFlux AmeriFlux US-FR2 Freeman Ranch- Mesquite Juniper, https://doi.org/10.17190/AMF/1246054, 2016.

Litvak, M.: AmeriFlux FLUXNET-1F US-Mpj Mountainair Pinyon-Juniper Woodland, https://doi.org/10.17190/AMF/1832161, 2021.

Litvak, M.: AmeriFlux FLUXNET-1F US-Wjs Willard Juniper Savannah, https://doi.org/10.17190/AMF/1871146, 2022.

Litvak, M.: AmeriFlux FLUXNET-1F US-Seg Sevilleta grassland, https://doi.org/10.17190/AMF/1984572, 2023a.

Litvak, M.: AmeriFlux FLUXNET-1F US-Ses Sevilleta shrubland, https://doi.org/10.17190/AMF/1984573, 2023b.

Litvak, M.: AmeriFlux FLUXNET-1F US-Vcm Valles Caldera Mixed Conifer, https://doi.org/10.17190/AMF/2229391, 2023c.

Litvak, M.: AmeriFlux FLUXNET-1F US-Vcp Valles Caldera Ponderosa Pine, https://doi.org/10.17190/AMF/2229392, 2023d.

Liu, L., Zhou, W., Guan, K., Peng, B., Xu, S., Tang, J., Zhu, Q., Till, J., Jia, X., Jiang, C., Wang, S., Qin, Z., Kong, H., Grant, R., Mezbahuddin, S., Kumar, V., and Jin, Z.: Knowledge-guided machine learning can improve carbon cycle quantification in agroecosystems, Nat Commun, 15, 357, https://doi.org/10.1038/s41467-023-43860-5, 2024.

Liu, Y., Holtzman, N. M., and Konings, A. G.: Global ecosystem-scale plant hydraulic traits retrieved using model–data fusion, Hydrology and Earth System Sciences, 25, 2399–2417, https://doi.org/10.5194/hess-25-2399-2021, 2021.

Lohila, A., Aurela, M., Tuovinen, J.-P., Hatakka, J., and Laurila, T.: FLUXNET2015 FI-Jok Jokioinen, https://doi.org/10.18140/FLX/1440159, 2016.

Lund, M., Jackowicz-Korczyński, M., and Abermann, J.: FLUXNET2015 GL-ZaF Zackenberg Fen, https://doi.org/10.18140/FLX/1440223, 2016a.

Lund, M., Jackowicz-Korczyński, M., and Abermann, J.: FLUXNET2015 GL-ZaH Zackenberg Heath, https://doi.org/10.18140/FLX/1440224, 2016b.

Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, in: Advances in Neural Information Processing Systems, 2017.

Luo, X., Zhou, H., Satriawan, T. W., Tian, J., Zhao, R., Keenan, T. F., Griffith, D. M., Sitch, S., Smith, N. G., and Still, C. J.: Mapping the global distribution of C4 vegetation using observations and optimality theory, Nat Commun, 15, 1219, https://doi.org/10.1038/s41467-024-45606-3, 2024.

Ma, H. and Liang, S.: Development of the GLASS 250-m leaf area index product (version 6) from MODIS data using the bidirectional LSTM deep learning model, Remote Sensing of Environment, 273, 112985, https://doi.org/10.1016/j.rse.2022.112985, 2022.

Ma, H., Zeng, J., Chen, N., Zhang, X., Cosh, M. H., and Wang, W.: Satellite surface soil moisture from SMAP, SMOS, AMSR2 and ESA CCI: A comprehensive assessment using global ground-based observations, Remote Sensing of Environment, 231, 111215, https://doi.org/10.1016/j.rse.2019.111215, 2019.

Ma, Y., Zhang, Z., Kang, Y., and Özdoğan, M.: Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach, Remote Sensing of Environment, 259, 112408, https://doi.org/10.1016/j.rse.2021.112408, 2021.

Macfarlane, C., Lambert, P., Byrne, J., Johnstone, C., and Smart, N.: FLUXNET2015 AU-Gin Gingin, https://doi.org/10.18140/FLX/1440199, 2016.

Manca, G. and Goded, I.: FLUXNET2015 IT-PT1 Parco Ticino forest, https://doi.org/10.18140/FLX/1440172, 2016.

Margolis, H.: AmeriFlux AmeriFlux CA-Qc2 Quebec - 1975 Harvested Black Spruce (HBS75), https://doi.org/10.17190/AMF/1419514, 2018.

Margolis, H. A.: AmeriFlux FLUXNET-1F CA-Qfo Quebec - Eastern Boreal, Mature Black Spruce, https://doi.org/10.17190/AMF/2006960, 2023.

Massman, B.: FLUXNET2015 US-GBT GLEES Brooklyn Tower, https://doi.org/10.18140/FLX/1440118, 2016a.

Massman, B.: FLUXNET2015 US-GLE GLEES, https://doi.org/10.18140/FLX/1440069, 2016b.

Matteucci, G.: FLUXNET2015 IT-Col Collelongo, https://doi.org/10.18140/FLX/1440167, 2016.

McCaughey, H.: AmeriFlux FLUXNET-1F CA-Gro Ontario - Groundhog River, Boreal Mixedwood Forest, https://doi.org/10.17190/AMF/1902823, 2022.

Merbold, L., Rebmann, C., and Corradi, C.: FLUXNET2015 RU-Che Cherski, https://doi.org/10.18140/FLX/1440181, 2016.

Meyer, W., Cale, P., Koerber, G., Ewenz, C., and Sun, Q.: FLUXNET2015 AU-Cpr Calperum, https://doi.org/10.18140/FLX/1440195, 2016.

Meyers, T.: FLUXNET2015 US-Goo Goodwin Creek, https://doi.org/10.18140/FLX/1440070, 2016.

Munger, J. W.: FLUXNET2015 US-Ha1 Harvard Forest EMS Tower (HFR1), https://doi.org/10.18140/FLX/1440071, 2016.

Myneni, R., Knyazikhin, Y., and Park, T.: MCD15A3H MODIS/Terra+Aqua Leaf Area Index/FPAR 4-day L4 Global 500m SIN Grid V006 [Data set], NASA EOSDIS Land Processes DAAC., https://doi.org/10.5067/MODIS/MCD15A3H.006, 2015a.

Myneni, R., Knyazikhin, Y., and Park, T.: MOD15A2H MODIS/Terra Leaf Area Index/FPAR 8-Day L4 Global 500m SIN Grid V006 [Data set], NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/MODIS/MOD15A2H.006, 2015b.

Nelson, J. A., Walther, S., Gans, F., Kraft, B., Weber, U., Novick, K., Buchmann, N., Migliavacca, M., Wohlfahrt, G., Šigut, L., Ibrom, A., Papale, D., Göckede, M., Duveiller, G., Knohl, A., Hörtnagl, L., Scott, R. L., Zhang, W., Hamdi, Z. M., Reichstein, M., Aranda-Barranco, S., Ardö, J., Op de Beeck, M., Billdesbach, D., Bowling, D., Bracho, R., Brümmer, C., Camps-Valls, G., Chen, S., Cleverly, J. R., Desai, A., Dong, G., El-Madany, T. S., Euskirchen, E. S., Feigenwinter, I., Galvagno, M., Gerosa, G., Gielen, B., Goded, I., Goslee, S., Gough, C. M., Heinesch, B., Ichii, K., Jackowicz-Korczynski, M. A., Klosterhalfen, A., Knox, S., Kobayashi, H., Kohonen, K.-M., Korkiakoski, M., Mammarella, I., Mana, G., Marzuoli, R., Matamala, R., Metzger, S., Montagnani, L., Nicolini, G., O'Halloran, T., Ourcival, J.-M., Peichl, M., Pendall, E., Ruiz Reverter, B., Roland, M., Sabbatini, S., Sachs, T., Schmidt, M., Schwalm, C. R., Shekhar, A., Silberstein, R., Silveira, M. L., Spano, D., Tagesson, T., Tramontana, G., Trotta, C., Turco, F., Vesala, T., Vincke, C., Vitale, D., Vivoni, E. R., Wang, Y., Woodgate, W., Yepez, E. A., Zhang, J., Zona, D., and Jung, M.: X-BASE: the first terrestrial carbon and water flux products from an extended data-driven scaling framework, FLUXCOM-X, EGUsphere, 1–51, https://doi.org/10.5194/egusphere-2024-165, 2024.

Nouvellon, Y.: FLUXNET2015 CG-Tch Tchizalamou, https://doi.org/10.18140/FLX/1440142, 2016.

Novick, K. and Phillips, R.: AmeriFlux FLUXNET-1F US-MMS Morgan Monroe State Forest, https://doi.org/10.17190/AMF/1854369, 2022.

Oishi, C., Novick, K., and Stoy, P.: AmeriFlux AmeriFlux US-Dk1 Duke Forest-open field, https://doi.org/10.17190/AMF/1246046, 2016a.

Oishi, C., Novick, K., and Stoy, P.: AmeriFlux AmeriFlux US-Dk2 Duke Forest-hardwoods, https://doi.org/10.17190/AMF/1246047, 2016b.

Oishi, C., Novick, K., and Stoy, P.: AmeriFlux AmeriFlux US-Dk3 Duke Forest - loblolly pine, https://doi.org/10.17190/AMF/1246048, 2016c.

O'Sullivan, M., Spracklen, D. V., Batterman, S. A., Arnold, S. R., Gloor, M., and Buermann, W.: Have Synergies Between Nitrogen Deposition and Atmospheric CO 2 Driven the Recent Enhancement of the Terrestrial Carbon Sink?, Global Biogeochemical Cycles, 33, 163–180, https://doi.org/10.1029/2018GB005922, 2019.

O'Sullivan, M., Smith, W. K., Sitch, S., Friedlingstein, P., Arora, V. K., Haverd, V., Jain, A. K., Kato, E., Kautz, M., Lombardozzi, D., Nabel, J. E. M. S., Tian, H., Vuichard, N., Wiltshire, A., Zhu, D., and Buermann, W.: Climate-Driven Variability and Trends in Plant Productivity Over Recent Decades Based on Three Global Products, Global Biogeochemical Cycles, 34, https://doi.org/10.1029/2020GB006613, 2020.

Ourcival, J.-M., Piquemal, K., Joffre, R., and Jean-Marc, L.: FLUXNET2015 FR-Pue Puechabon, https://doi.org/10.18140/FLX/1440164, 2016.

Ozflux, P.: FluxNet Data OzFlux: Australian and New Zealand Flux Research and Monitoring, https://data.ozflux.org.au/portal/pub/viewColDetails.jspx?collection.id=1882723&collection.owner.id=450&viewType=anonymous, 2024.

Papale, D., Tirone, G., Valentini, R., Arriga, N., Belelli, L., Consalvo, C., Dore, S., Manca, G., Mazzenga, F., Sabbatini, S., Stefani, P., Boschi, A., and Tomassucci, M.: FLUXNET2015 IT-Ro2 Roccarespampani 2, https://doi.org/10.18140/FLX/1440175, 2016.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y. W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Ribeca, A., van Ingen, C., Zhang, L., Amiro, B., Ammann, C., Arain, M. A., Ardö, J., Arkebauer, T., Arndt, S. K., Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L. B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Boike, J., Bolstad, P. V., Bonal, D., Bonnefond, J. M., Bowling, D. R., Bracho, R., Brodeur, J., Brümmer, C., Buchmann, N., Burban, B., Burns, S. P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T. R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B. D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P. S., D'Andrea, E., da Rocha, H., Dai, X., Davis, K. J., De Cinti, B., de Grandcourt, A., De

Ligne, A., De Oliveira, R. C., Delpierre, N., Desai, A. R., Di Bella, C. M., di Tommasi, P., Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrêne, E., Dunn, A., Dušek, J., Eamus, D., Eichelmann, U., ElKhidir, H. A. M., Eugster, W., Ewenz, C. M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., Gharun, M., Gianelle, D., et al.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Scientific data, 7, 225, https://doi.org/10.1038/s41597-020-0534-3, 2020.

Pendall, E., Griebel, A., Barton, C., and Metzen, D.: FLUXNET2015 AU-Cum Cumberland Plains, https://doi.org/10.18140/FLX/1440196, 2016.

Peñuelas, J., Ciais, P., Canadell, J. G., Janssens, I. A., Fernández-Martínez, M., Carnicer, J., Obersteiner, M., Piao, S., Vautard, R., and Sardans, J.: Shifting from a fertilization-dominated to a warming-dominated period, Nature Ecology and Evolution, 1, 1438–1445, https://doi.org/10.1038/s41559-017-0274-8, 2017.

Piao, S., Wang, X., Park, T., Chen, C., Lian, X., He, Y., Bjerke, J. W., Chen, A., Ciais, P., Tømmervik, H., Nemani, R. R., and Myneni, R. B.: Characteristics, drivers and feedbacks of global greening, Nature Reviews Earth and Environment, 1, 14–27, https://doi.org/10.1038/s43017-019-0001-x, 2020.

Pilegaard, K. and Ibrom, A.: FLUXNET2015 DK-Eng Enghave, https://doi.org/10.18140/FLX/1440153, 2016.

Posse, G., Lewczuk, N., Richter, K., and Cristiano, P.: FLUXNET2015 AR-Vir Virasoro, FluxNet; Instituto Nacional de Tecnología Agropecuaria, https://doi.org/10.18140/FLX/1440192, 2016.

Poveda, F. D., Ballesteros, A. L., Cañete, E. P. S., Ortiz, P. S., Jiménez, M. R. M., Priego, O. P., and Kowalski, A. S.: FLUXNET2015 ES-Amo Amoladeras, https://doi.org/10.18140/FLX/1440156, 2016.

Prentice, I. C., Dong, N., Gleason, S. M., Maire, V., and Wright, I. J.: Balancing the costs of carbon gain and water transport: testing a new theoretical framework for plant functional ecology, Ecology Letters, 17, 82–91, https://doi.org/10.1111/ele.12211, 2014.

Reich, P. B., Hobbie, S. E., and Lee, T. D.: Plant growth enhancement by elevated CO2 eliminated by joint water and nitrogen limitation, Nature Geoscience, 7, 920–924, https://doi.org/10.1038/ngeo2284, 2014.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, https://doi.org/10.1038/s41586-019-0912-1, 2019.

Reverter, B. R., Perez-Cañete, E. S., and Kowalski, A. S.: FLUXNET2015 ES-LgS Laguna Seca, https://doi.org/10.18140/FLX/1440225, 2016.

Richardson, A. and Hollinger, D.: AmeriFlux FLUXNET-1F US-Bar Bartlett Experimental Forest, https://doi.org/10.17190/AMF/2006969, 2023.

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, Nat Rev Earth Environ, 1–17, https://doi.org/10.1038/s43017-023-00456-3, 2023.

Running, S., Mu, Q., and Zhao, M.: MOD17A2H MODIS/Terra Gross Primary Productivity 8-Day L4 Global 500m SIN Grid V006, https://doi.org/10.5067/MODIS/MOD17A2H.006, 2015.

Ryu, Y., Jiang, C., Kobayashi, H., and Detto, M.: MODIS-derived global land products of shortwave radiation and diffuse and total photosynthetically active radiation at 5 km resolution from 2000, Remote Sensing of Environment, 204, 812–825, https://doi.org/10.1016/j.rse.2017.09.021, 2018.

Ryu, Y., Berry, J. A., and Baldocchi, D. D.: What is global photosynthesis? History, uncertainties and opportunities, Remote Sensing of Environment, 223, 95–114, https://doi.org/10.1016/j.rse.2019.01.016, 2019.

Sabater, J. M.: ERA5-Land monthly averaged data from 1981 to present., Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/doi:10.24381/cds.68d2bb30, 2019.

Sabbatini, S., Arriga, N., Papale, D., Boschi, A., and Tomassucci, M.: FLUXNET2015 IT-CA1 Castel d'Asso1, https://doi.org/10.18140/FLX/1440230, 2016a.

Sabbatini, S., Arriga, N., Gioli, B., Papale, D., Boschi, A., and Tomassucci, M.: FLUXNET2015 IT-CA2 Castel d'Asso2, https://doi.org/10.18140/FLX/1440231, 2016b.

Sabbatini, S., Arriga, N., Matteucci, G., Papale, D., Boschi, A., and Tomassucci, M.: FLUXNET2015 IT-CA3 Castel d'Asso 3, https://doi.org/10.18140/FLX/1440232, 2016c.

Saleska, S.: FLUXNET2015 BR-Sa1 Santarem-Km67-Primary Forest, https://doi.org/10.18140/FLX/1440032, 2016.

Schaaf, C. and Wang, Z.: MCD43C4 MODIS/Terra+Aqua BRDF/Albedo Nadir BRDF-Adjusted Ref Daily L3 Global 0.05Deg CMG V006 [Data set], NASA EOSDIS Land Processes DAAC., https://doi.org/10.5067/MODIS/MCD43C4.006, 2015.

Schneider, K. and Schmidt, M.: FLUXNET2015 DE-Seh Selhausen, https://doi.org/10.18140/FLX/1440217, 2016.

Schroder, I., Zegelin, S., Palu, T., and Feitz, A.: FLUXNET2015 AU-Emr Emerald, https://doi.org/10.18140/FLX/1440198, 2016.

Schwalm, C. R., Anderegg, W. R. L., Michalak, A. M., Fisher, J. B., Biondi, F., Koch, G., Litvak, M., Ogle, K., Shaw, J. D., Wolf, A., Huntzinger, D. N., Schaefer, K., Cook, R., Wei, Y., Fang, Y., Hayes, D., Huang, M., Jain, A., and Tian, H.: Global patterns of drought recovery, Nature, 548, 202–205, https://doi.org/10.1038/nature23021, 2017.

Scott, R.: FLUXNET2015 US-SRM Santa Rita Mesquite, https://doi.org/10.18140/FLX/1440090, 2016a.

Scott, R.: FLUXNET2015 US-Whs Walnut Gulch Lucky Hills Shrub, https://doi.org/10.18140/FLX/1440097, 2016b.

Shao, C.: FLUXNET2015 CN-Sw2 Siziwang Grazed (SZWG), https://doi.org/10.18140/FLX/1440212, 2016.

Sigut, L., Havrankova, K., Jocher, G., Pavelka, M., Janouš, D., Czerny, R., Stanik, K., and Trusina, J.: FLUXNET2015 CZ-BK2 Bily Kriz grassland, https://doi.org/10.18140/FLX/1440144, 2016.

Smith, W. K., Reed, S. C., Cleveland, C. C., Ballantyne, A. P., Anderegg, W. R. L., Wieder, W. R., Liu, Y. Y., and Running, S. W.: Large divergence of satellite and Earth system model estimates of global terrestrial CO2 fertilization, Nature Climate Change, 6, 306–310, https://doi.org/10.1038/nclimate2879, 2016.

Spano, D., Duce, P., Marras, S., Sirca, C., Arca, A., Zara, P., Ventura, A., Mereu, S., and Sanna, L.: FLUXNET2015 IT-Noe Arca di Noe - Le Prigionette, https://doi.org/10.18140/FLX/1440171, 2016.

Staebler, R.: AmeriFlux FLUXNET-1F CA-Cbo Ontario - Mixed Deciduous, Borden Forest Site, https://doi.org/10.17190/AMF/1854365, 2022.

Ștefan, V. and Levin, S.: plotbiomes: R package for plotting Whittaker biomes with ggplot2, , https://doi.org/10.5281/zenodo.7145245, 2018.

Still, C. J., Berry, J. A., Collatz, G. J., and DeFries, R. S.: Global distribution of C3 and C4 vegetation: Carbon cycle implications, Global Biogeochemical Cycles, 17, https://doi.org/10.1029/2001gb001807, 2003.

Still, C. J., Berry, J. A., Collatz, G. J., and DeFries, R. S.: ISLSCP II C4 Vegetation Percentage, in: Hall, Forrest G., G. Collatz, B. Meeson, S. Los, E. Brown de Colstoun, and D. Landis (eds.). ISLSCP Initiative II Collection. Data set., http://dx.doi.org/10.3334/ORNLDAAC/932, 2009.

Stocker, B. D., Zscheischler, J., Keenan, T. F., Prentice, I. C., Peñuelas, J., and Seneviratne, S. I.: Quantifying soil moisture impacts on light use efficiency across biomes, New Phytologist, 218, 1430–1449, https://doi.org/10.1111/nph.15123, 2018.

Stocker, B. D., Zscheischler, J., Keenan, T. F., Prentice, I. C., Seneviratne, S. I., and Peñuelas, J.: Drought impacts on terrestrial primary production underestimated by satellite monitoring, Nature Geoscience, 12, 264–270, https://doi.org/10.1038/s41561-019-0318-6, 2019.

Stocker, B. D., Tumber-Dávila, S. J., Konings, A. G., Anderson, M. C., Hain, C., and Jackson, R. B.: Global patterns of water storage in the rooting zones of vegetation, Nat. Geosci., 16, 250–256, https://doi.org/10.1038/s41561-023-01125-2, 2023.

Sturtevant, C., Szutu, D., Baldocchi, D., Matthes, J. H., Oikawa, P., and Chamberlain, S. D.: FLUXNET2015 US-Myb Mayberry Wetland, https://doi.org/10.18140/FLX/1440105, 2016.

Suyker, A.: FLUXNET2015 US-Ne1 Mead - irrigated continuous maize site, https://doi.org/10.18140/FLX/1440084, 2016a.

Suyker, A.: FLUXNET2015 US-Ne2 Mead - irrigated maize-soybean rotation site, https://doi.org/10.18140/FLX/1440085, 2016b.

Suyker, A.: FLUXNET2015 US-Ne3 Mead - rainfed maize-soybean rotation site, https://doi.org/10.18140/FLX/1440086, 2016c.

Tagesson, T., Ardö, J., and Fensholt, R.: FLUXNET2015 SN-Dhr Dahra, https://doi.org/10.18140/FLX/1440246, 2016.

Tang, Y., Kato, T., and Du, M.: FLUXNET2015 CN-HaM Haibei Alpine Tibet site, https://doi.org/10.18140/FLX/1440190, 2016.

Terrer, C., Jackson, R. B., Prentice, I. C., Keenan, T. F., Kaiser, C., Vicca, S., Fisher, J. B., Reich, P. B., Stocker, B. D., Hungate, B. A., Peñuelas, J., McCallum, I., Soudzilovskaia, N. A., Cernusak, L. A., Talhelm, A. F., Van Sundert, K., Piao, S., Newton, P. C. D., Hovenden, M. J., Blumenthal, D. M., Liu, Y. Y., Müller, C., Winter, K., Field, C. B., Viechtbauer, W., Van Lissa, C. J., Hoosbeek, M. R., Watanabe, M., Koike, T., Leshyk, V. O., Polley, H. W., and Franklin, O.: Nitrogen and phosphorus constrain the CO2 fertilization of global plant biomass, Nature Climate Change, 9, 684–689, https://doi.org/10.1038/s41558-019-0545-2, 2019.

Thoning, K. W., Crotwell, A. M., and Mund, J. W.: Atmospheric Carbon Dioxide Dry Air Mole Fractions from continuous measurements at Mauna Loa, Hawaii, Barrow, Alaska, American Samoa and South Pole. 1973-2020, Version 2021-08-09, National Oceanic and Atmospheric Administration (NOAA), Global Monitoring Laboratory (GML), Boulder, Colorado, USA, https://doi.org/10.15138/yaf1-bk21, 2021.

Tramontana, G., Ichii, K., Camps-Valls, G., Tomelleri, E., and Papale, D.: Uncertainty analysis of gross primary production upscaling using Random Forests, remote sensing and eddy covariance data, Remote Sensing of Environment, 168, 360–373, https://doi.org/10.1016/j.rse.2015.07.015, 2015.

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, 2016.

Ueyama, M., Iwata, H., and Harazono, Y.: AmeriFlux AmeriFlux US-Uaf University of Alaska, Fairbanks, https://doi.org/10.17190/AMF/1480322, 2018.

Ueyama, M., Ichii, K., Kobayashi, H., Kumagai, T., Beringer, J., Merbold, L., Euskirchen, E. S., Hirano, T., Marchesini, L. B., Baldocchi, D., Saitoh, T. M., Mizoguchi, Y., Ono, K., Kim, J., Varlagin, A., Kang, M., Shimizu, T., Kosugi, Y., Bret-Harte, M. S., Machimura, T., Matsuura, Y., Ohta, T., Takagi, K., Takanashi, S., and Yasuda, Y.: Inferring CO2 fertilization effect based on global monitoring land-atmosphere exchange with a theoretical model, Environmental Research Letters, 15, 84009, https://doi.org/10.1088/1748-9326/ab79e5, 2020.

Valach, A., Shortt, R., Szutu, D., Eichelmann, E., Knox, S., Hemes, K., Verfaillie, J., and Baldocchi, D.: AmeriFlux FLUXNET-1F US-Tw1 Twitchell Wetland West Pond, https://doi.org/10.17190/AMF/1832165, 2021.

Valentini, R., Nicolini, G., Stefani, P., de Grandcourt, A., and Stivanello, S.: FLUXNET2015 GH-Ank Ankasa, https://doi.org/10.18140/FLX/1440229, 2016a.

Valentini, R., Dore, S., Mazzenga, F., Sabbatini, S., Stefani, P., Tirone, G., and Papale, D.: FLUXNET2015 IT-Cpz Castelporziano, https://doi.org/10.18140/FLX/1440168, 2016b.

Valentini, R., Tirone, G., Vitale, D., Papale, D., Arriga, N., Belelli, L., Dore, S., Manca, G., Mazzenga, F., Pegoraro, E., Sabbatini, S., Stefani, P., Boschi, A., and Tomassucci, M.: FLUXNET2015 IT-Ro1 Roccarespampani 1, https://doi.org/10.18140/FLX/1440174, 2016c.

Villarreal, S. and Vargas, R.: Representativeness of FLUXNET Sites Across Latin America, Journal of Geophysical Research: Biogeosciences, 126, e2020JG006090, https://doi.org/10.1029/2020JG006090, 2021.

Walker, A. P., De Kauwe, M. G., Bastos, A., Belmecheri, S., Georgiou, K., Keeling, R. F., McMahon, S. M., Medlyn, B. E., Moore, D. J. P., Norby, R. J., Zaehle, S., Anderson-Teixeira, K. J., Battipaglia, G., Brienen, R. J. W., Cabugao, K. G., Cailleret, M., Campbell, E., Canadell, J. G., Ciais, P., Craig, M. E., Ellsworth, D. S., Farquhar, G. D., Fatichi, S., Fisher, J. B., Frank, D. C., Graven, H., Gu, L., Haverd, V., Heilman, K., Heimann, M., Hungate, B. A., Iversen, C. M., Joos, F., Jiang, M., Keenan, T. F., Knauer, J., Körner, C., Leshyk, V. O., Leuzinger, S., Liu, Y., MacBean, N., Malhi, Y., McVicar, T. R., Penuelas, J., Pongratz, J., Powell, A. S., Riutta, T., Sabot, M. E. B., Schleucher, J., Sitch, S., Smith, W. K., Sulman, B., Taylor, B., Terrer, C., Torn, M. S., Treseder, K. K., Trugman, A. T., Trumbore, S. E., van Mantgem, P. J., Voelker, S. L., Whelan, M. E., and Zuidema, P. A.: Integrating the evidence for a terrestrial carbon sink caused by increasing atmospheric CO2, New Phytologist, 229, 2413–2445, https://doi.org/10.1111/nph.16866, 2021.

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

Wan, Z., Hook, S., and Hulley, G.: MOD11A1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006 [Data set], NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/MODIS/MOD11A1.006, 2015a.

Wan, Z., Hook, S., and Hulley, G.: MYD11A1 MODIS/Aqua Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006 [Data set], NASA EOSDIS Land Processes DAAC, https://doi.org/10.5067/MODIS/MYD11A1.006, 2015b.

Wang, H. and Fu, X.: FLUXNET2015 CN-Qia Qianyanzhou, https://doi.org/10.18140/FLX/1440141, 2016.

Wang, H., Prentice, I. C., Keenan, T. F., Davis, T. W., Wright, I. J., Cornwell, W. K., Evans, B. J., and Peng, C.: Towards a universal model for carbon dioxide uptake by plants, Nature Plants, 3, 734–741, https://doi.org/10.1038/s41477-017-0006-8, 2017.

Warm Winter 2020 Team: Warm Winter 2020 ecosystem eddy covariance flux product for 73 stations in FLUXNET-Archive format—release 2022-1 (Version 1.0), https://doi.org/10.18160/2G60-ZHAK, 2022.

Wenzel, S., Cox, P. M., Eyring, V., and Friedlingstein, P.: Projected land photosynthesis constrained by changes in the seasonal cycle of atmospheric CO2, Nature, 538, 499–501, https://doi.org/10.1038/nature19772, 2016.

Wohlfahrt, G., Hammerle, A., and Hörtnagl, L.: FLUXNET2015 AT-Neu Neustift, https://doi.org/10.18140/FLX/1440121, 2016.

Wolf, S., Eugster, W., and Buchmann, N.: FLUXNET2015 PA-SPn Sardinilla Plantation, https://doi.org/10.18140/FLX/1440180, 2016.

Woodgate, W., van Gorsel, E., Leuning, R., Hughes, D., Kitchen, M., and Zegelin, S.: FLUXNET2015 AU-Tum Tumbarumba, https://doi.org/10.18140/FLX/1440126, 2016.

Xiao, J., Zhuang, Q., Baldocchi, D. D., Law, B. E., Richardson, A. D., Chen, J., Oren, R., Starr, G., Noormets, A., Ma, S., Verma, S. B., Wharton, S., Wofsy, S. C., Bolstad, P. V., Burns, S. P., Cook, D. R., Curtis, P. S., Drake, B. G., Falk, M., Fischer, M. L., Foster, D. R., Gu, L., Hadley, J. L., Hollinger, D. Y., Katul, G. G., Litvak, M., Martin, T. A., Matamala, R., McNulty, S., Meyers, T. P., Monson, R. K., Munger, J. W., Oechel, W. C., Paw U, K. T., Schmid, H. P., Scott, R. L., Sun, G., Suyker, A. E., and Torn, M. S.: Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data, Agricultural and Forest Meteorology, 148, 1827–1847, https://doi.org/10.1016/j.agrformet.2008.06.015, 2008.

Xie, X., Chen, J. M., Yuan, W., Guan, X., Jin, H., and Leng, J.: A Practical Algorithm for Correcting Topographical Effects on Global GPP Products, Journal of Geophysical Research: Biogeosciences, 128, e2023JG007553, https://doi.org/10.1029/2023JG007553, 2023.

Xie, Y., Gibbs, H. K., and Lark, T. J.: Landsat-based Irrigation Dataset (LANID): 30m resolution maps of irrigation distribution, frequency, and change for the US, 1997–2017, Earth System Science Data, 13, 5689–5710, https://doi.org/10.5194/essd-13-5689-2021, 2021.

Yan, K., Park, T., Yan, G., Chen, C., Yang, B., Liu, Z., Nemani, R. R., Knyazikhin, Y., and Myneni, R. B.: Evaluation of MODIS LAI/FPAR product collection 6. Part 1: Consistency and improvements, Remote Sensing, 8, 1–16, https://doi.org/10.3390/rs8050359, 2016a.

Yan, K., Park, T., Yan, G., Liu, Z., Yang, B., Chen, C., Nemani, R. R., Knyazikhin, Y., and Myneni, R. B.: Evaluation of MODIS LAI/FPAR product collection 6. Part 2: Validation and intercomparison, Remote Sensing, 8, 460, https://doi.org/10.3390/rs8060460, 2016b.

Yang, F., Ichii, K., White, M. A., Hashimoto, H., Michaelis, A. R., Votava, P., Zhu, A. X., Huete, A., Running, S. W., and Nemani, R. R.: Developing a continental-scale measure of gross primary production by combining MODIS and AmeriFlux data through Support Vector Machine approach, Remote Sensing of Environment, 110, 109–122, https://doi.org/10.1016/j.rse.2007.02.016, 2007.

Yang, R., Wang, J., Zeng, N., Sitch, S., Tang, W., McGrath, M. J., Cai, Q., Liu, D., Lombardozzi, D., Tian, H., Jain, A. K., and Han, P.: Divergent historical GPP trends among state-of-the-art multi-model simulations and satellite-based products, Earth System Dynamics, 13, 833–849, https://doi.org/10.5194/esd-13-833-2022, 2022.

Yuan, H., Dai, Y., Xiao, Z., Ji, D., and Shangguan, W.: Reprocessing the MODIS Leaf Area Index products for land surface and climate modelling, Remote Sensing of Environment, 115, 1171–1187, https://doi.org/10.1016/j.rse.2011.01.001, 2011.

Zeng, J., Matsunaga, T., Tan, Z.-H., Saigusa, N., Shirai, T., Tang, Y., Peng, S., and Fukuda, Y.: Global terrestrial carbon fluxes of 1999–2019 estimated by upscaling eddy covariance data with a random forest, Sci Data, 7, 313, https://doi.org/10.1038/s41597-020-00653-5, 2020.

Zeng, N., Zhao, F., Collatz, G. J., Kalnay, E., Salawitch, R. J., West, T. O., and Guanter, L.: Agricultural Green Revolution as a driver of increasing atmospheric CO2 seasonal amplitude, Nature, 515, 394–397, https://doi.org/10.1038/nature13893, 2014.

Zhan, C., Orth, R., Migliavacca, M., Zaehle, S., Reichstein, M., Engel, J., Rammig, A., and Winkler, A. J.: Emergence of the physiological effects of elevated CO2 on land–atmosphere exchange of carbon and water, Global Change Biology, 28, 7313–7326, https://doi.org/10.1111/gcb.16397, 2022.

Zhang, J. and Han, S.: FLUXNET2015 CN-Cha Changbaishan, https://doi.org/10.18140/FLX/1440137, 2016.

Zhang, Y.: A global spatially contiguous solar-induced fluorescence (CSIF) dataset using neural networks (2000-2020), National Tibetan Plateau Data Center, https://doi.org/10.11888/Ecolo.tpdc.271751., 2021.

Zhang, Y., Joiner, J., Hamed Alemohammad, S., Zhou, S., and Gentine, P.: A global spatially contiguous solar-induced fluorescence (CSIF) dataset using neural networks, Biogeosciences, 15, 5779–5800, https://doi.org/10.5194/bg-15-5779-2018, 2018.

Zhang, Y., Kong, D., Gan, R., Chiew, F. H. S., McVicar, T. R., Zhang, Q., and Yang, Y.: Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017, Remote Sensing of Environment, 222, 165–182, https://doi.org/10.1016/j.rse.2018.12.031, 2019.

Zheng, Y., Shen, R., Wang, Y., Li, X., Liu, S., Liang, S., Chen, J. M., Ju, W., Zhang, L., and Yuan, W.: Improved estimate of global gross primary production for reproducing its long-Term variation, 1982-2017, Earth System Science Data, 12, 2725–2746, https://doi.org/10.5194/essd-12-2725-2020, 2020.

Zhou, G. and Yan, J.: FLUXNET2015 CN-Din Dinghushan, https://doi.org/10.18140/FLX/1440139, 2016.

Zhu, Z., Piao, S., Myneni, R. B., Huang, M., Zeng, Z., Canadell, J. G., Ciais, P., Sitch, S., Friedlingstein, P., Arneth, A., Cao, C., Cheng, L., Kato, E., Koven, C., Li, Y., Lian, X., Liu, Y., Liu, R., Mao, J., Pan, Y., Peng, S., Peuelas, J., Poulter, B., Pugh, T. A. M., Stocker, B. D., Viovy, N., Wang, X., Wang, Y., Xiao, Z., Yang, H., Zaehle, S., and Zeng, N.: Greening of the Earth and its drivers, Nature Climate Change, 6, 791–795, https://doi.org/10.1038/nclimate3004, 2016.

Zhuang, J., dussin, raphael, Huard, D., Bourgault, P., Banihirwe, A., Raynaud, S., Malevich, B., Schupfner, M., Filipe, Levang, S., Gauthier, C., Jüling, A., Almansi, M., RichardScottOZ, RondeauG, Rasp, S., Smith, T. J., Stachelek, J., Plough, M., Pierre, Bell, R., Caneill, R., and Li, X.: xESMF: v0.8.2, , https://doi.org/10.5281/zenodo.8356796, 2023.

Zona, D. and Oechel, W.: FLUXNET2015 US-Atq Atqasuk, https://doi.org/10.18140/FLX/1440067, 2016a.

Zona, D. and Oechel, W.: FLUXNET2015 US-Ivo Ivotuk, https://doi.org/10.18140/FLX/1440073, 2016b.