ENVIRONMENTAL RESEARCH
ECOLOGY

PURPOSE-LED PUBLISHING™

PAPER • OPEN ACCESS

Using satellite-based Sun-induced chlorophyll fluorescence and spectral reflectance for improving terrestrial CO_2 flux estimates of India

To cite this article: Aparnna Ravi *et al* 2025 *Environ. Res.: Ecology* **4** 015004

ENVIRONMENTAL RESEARCH
ECOLOGY

OPEN ACCESS

RECEIVED
9 September 2024

REVISED
6 January 2025

ACCEPTED FOR PUBLICATION
20 January 2025

PUBLISHED
29 January 2025

PAPER

Using satellite-based Sun-induced chlorophyll fluorescence and spectral reflectance for improving terrestrial CO_2 flux estimates of India

Aparnna Ravi[1,2], **Dhanyalekshmi Pillai**[1,2,*], **Christoph Gerbig**[3], **Stephen Sitch**[4], **Sönke Zaehle**[3], **Vishnu Thilakan**[1,2,6], **Chandra Sekhar Jha**[5] **and Thara Anna Mathew**[1,2]

[1] Indian Institute of Science Education and Research Bhopal (IISERB), Bhopal, India
[2] Max Planck Partner Group at IISERB, Bhopal, India
[3] Max-Planck Institute of Biogeochemistry, Jena, Germany
[4] University of Exeter, Exeter EX4 4QF, United Kingdom
[5] National Remote Sensing Centre (ISRO), Balanagar, Hyderabad, India
[6] Present address: Lund University, Lund, Sweden
[*] Author to whom any correspondence should be addressed.

E-mail: dhanya@iiserb.ac.in

Keywords: net ecosystem exchange, VPRM, SIF, MODIS, eddy covariance, TROPOMI, OCO-2

Supplementary material for this article is available online

Abstract

Significant uncertainties in terrestrial carbon fluxes exist in regions with limited ground-based observations, impacting our understanding of ecosystem carbon dynamics and emission reduction needs. This is particularly true for areas with sparse measurement networks, like India. To address this, we explore the potential of satellite measurements from various missions such as Sentinel-5 Precursor and the Orbiting Carbon Observatory-2 to improve terrestrial biosphere CO_2 fluxes of India. We follow a data-driven approach, which simulates spatial and temporal distributions of gross primary productivity (GPP), net ecosystem exchange (NEE), and ecosystem respiration (R_{eco}). We improve these model predictions by additionally using satellite-based solar-induced chlorophyll fluorescence (SIF), soil temperature , and soil moisture specific to the vegetation classes of the domain. Different model refinements were performed to present the improved hourly distributions of terrestrial biospheric CO_2 fluxes on a $0.1^\circ \times 0.1^\circ$ grid from 2012 to 2020. Among them, the best-performing model simulations show reasonable agreement with eddy covariance observations for 2012–2018. For example, our best NEE and GPP predictions are highly correlated with observations with squared correlation coefficient (R^2) values of 0.68 (NEE) and 0.74 (GPP) at the monthly scale for 2018. Based on our improved estimations, the annual NEE and GPP show values within the range from -0.38 Pg C yr^{-1} to -0.53 Pg C yr^{-1} (land C sink) and 3.39 Pg C yr^{-1} to 3.88 Pg C yr^{-1}, respectively over India for 2012–2020. Our novel approach and findings highlight the potential of satellite-based SIF measurements to detail the ecosystem-scale vegetation responses across various biomes in India. The use of satellite observations, as demonstrated in this study, offers a scalable solution for regions lacking sufficient ground-based observations to estimate biospheric carbon fluxes reliably.

1. Introduction

The terrestrial biosphere is the largest carbon sink, which sequesters about a quarter of global annual CO_2 emissions (Friedlingstein *et al* 2022). Major terrestrial biosphere carbon fluxes such as gross primary productivity (GPP), ecosystem respiration (R_{eco}), and net ecosystem exchange (NEE) exhibit significant spatiotemporal variability due to factors like vegetation type, age, climate response, geography, and other environmental conditions (van der Meer *et al* 2002). Hence, accurate estimation of these fluxes at finer scales

relevant to climate change mitigation remains challenging. Terrestrial biosphere models attempt to simulate these fluxes at global scales, yet uncertainties persist (e.g. Sitch *et al* 2008, 2015, Peylin *et al* 2013, Thompson *et al* 2016). Challenges for these models include limited global eddy covariance (EC) flux tower data for model validation or calibration, the incomplete representation of vital processes in the model (e.g. drought-related mortality), and insufficient understanding of environmental impacts on carbon exchanges. For instance, the existing models have limited observational data from India, affecting confidence in India's carbon flux estimates despite its global significance (Thompson *et al* 2016). Annual NEE estimates for India vary widely across studies (0 to -0.37 Pg C yr^{-1}) mirroring uncertainties in India's carbon flux estimation (Patra *et al* 2011, Nayak *et al* 2015, Sitch *et al* 2015, Rao *et al* 2019).

Atmospheric CO_2 measurements can be utilised in an atmospheric inverse modelling framework to evaluate and improve the terrestrial biosphere flux estimates of India. Careful selection of prior fluxes is crucial in solving the complex inverse problem (Rodgers 2000). Previous studies utilized biospheric models like the Carnegie Ames Stanford Approach (CASA; Potter *et al* 1993) and TRENDY model ensembles (Sitch *et al* 2015) to estimate biospheric CO_2 fluxes over southeast Asia and specifically India (Nayak *et al* 2010, 2013, Patra *et al* 2011, Peylin *et al* 2013, Goroshi *et al* 2014, Cervarich *et al* 2016). However, these models operate at coarse resolutions (typically 0.5° or coarser), with CASA providing monthly outputs and TRENDY offering sub-daily data (with output available monthly) but limited model validation against Indian observations. This coarse resolution limits the model's ability to capture detailed spatiotemporal flux distributions, resulting in varied estimates among studies (Patra *et al* 2013, Rao *et al* 2019, Friedlingstein *et al* 2022).

Recent satellite instruments measuring solar-induced chlorophyll fluorescence (SIF) from space are valuable for regions with limited ground-based observations. SIF retrievals capture re-emitted solar radiation (650–850 nm) from chlorophyll-a pigment, providing insights into carbon uptake through photosynthesis at regional to global scales (Frankenberg *et al* 2011, Sun *et al* 2017, Köhler *et al* 2018, Li *et al* 2018, Smith *et al* 2018, Gu *et al* 2019, Yu *et al* 2019, Zhang *et al* 2019, Liu *et al* 2023). Since the re-emission process (fluorescence) by chlorophyll is linked to the primary steps in photosynthesis, SIF can be used as the proxy for photosynthesis (Parazoo *et al* 2018, Sun *et al* 2018, Yu *et al* 2019), despite only about 2% of absorbed solar energy being re-emitted as fluorescence. So, SIF retrievals from space need advanced spectrometers with a high spectral resolution and a high signal-to-noise ratio due to narrow Fraunhofer lines and weak signals. It should be noted that SIF retrievals also are prone to errors related to signal strength, extraction range, leaf scattering, re-absorption effects, and background noise (Köhler *et al* 2015, Joiner *et al* 2016, Li *et al* 2018, Liu *et al* 2020). Limited available information makes it currently challenging to decouple those uncertainties from global retrievals.

The relationship between SIF and GPP can be linear or nonlinear, influenced by environmental variables like moisture, temperature, radiation, and precipitation, as well as observational factors such as wavelength and angle (Guanter *et al* 2012, 2014, Li *et al* 2018, Paul-Limoges *et al* 2018, Sun *et al* 2018, Wang *et al* 2020, Chen *et al* 2021, Kim *et al* 2021). Plant functional types and physiological factors further modulate the SIF-GPP relationship (Sun *et al* 2018). The first global SIF retrievals (at 755–775 nm) were done by the Greenhouse gases Observing SATellite (Frankenberg *et al* 2011). Other satellite missions providing SIF data at varying spatial and temporal resolutions include Global Ozone Monitoring Experiment 2 (Joiner *et al* 2013), Orbiting Carbon Observatory 2 (OCO-2; Sun *et al* 2018), OCO-3 (Taylor *et al* 2020), and TROPOspheric Monitoring Instrument (TROPOMI; Guanter *et al* 2021 onboard Sentinel-5 Precursor satellite).

This study presents the improved high-resolution mapping of terrestrial biosphere CO_2 fluxes across India from 2012 to 2020 at $0.1° \times 0.1°$ grid and hourly temporal resolution. This is achieved by employing a satellite-based data-driven approach using a biosphere model and combining it with surface reflectance, SIF, meteorological, and environmental variables to capture the spatiotemporal variations of biosphere fluxes. The gridded NEE, GPP, and R_{eco} are initially generated using a diagnostic satellite-based biosphere model, the Vegetation Photosynthesis and Respiration Model (VPRM; Mahadevan *et al* 2008). The VPRM model is driven by observational-based data, which utilizes model parameters optimized with direct EC measurements (Mahadevan *et al* 2008, Kountouris *et al* 2018, Gerbig and Koch 2021, Gourdji *et al* 2022, Bazzi *et al* 2024). Previously, Thilakan *et al* (2022) generated VPRM simulations ($0.1° \times 0.1°$, hourly) with uncalibrated model parameters. The novelty of the present study is the incorporation of the SIF-based information and soil-related variables in VPRM to improve the seasonality and magnitudes of NEE predictions in the absence of sufficient EC measurements to calibrate the model parameters. VPRM's benefit lies in its limited number of model parameters that need to be trained with direct measurements, making it ideal for carbon assimilation studies. These improved high-resolution biospheric flux products can be used in the near future as priors in the inverse data assimilation of CO_2 or can be coupled with high-resolution transport models for understanding the atmospheric CO_2 transport or variability associated with natural fluxes. The lack of reliable, observation-based biospheric carbon flux distribution at a resolution sufficient to

represent vegetation activities posed a significant challenge for the scientific community and policymakers, hindering effective climate change mitigation efforts in India. This study presents a novel approach by utilizing available observational proxies of vegetation activities, such as SIF and VIs from satellite remote sensing platforms, to address the limitations of ground-based observations and enable a comprehensive analysis of the biosphere–atmosphere carbon exchange in India.

2. Methods

For deriving improved estimates of terrestrial biosphere CO_2 fluxes across the ecosystems over India: we (i) implemented and customised the standard VPRM (hereafter referred to as $VPRM_{INI}$) for a domain covering India (5° N–40° N, 66° E–100° E, figure 2) and performed the simulations of NEE, GPP, and R_{eco} fluxes (section 2.1); (ii) derived ecosystem-specific linear relations between SIF and GPP and this information is provided to $VPRM_{INI}$ to improve the GPP estimates (section 2.2), and (iii) further modified the $VPRM_{INI}$-derived R_{eco} to include the influence of soil-related variables specific to each vegetation classes (section 2.3). The figure 1 provides an overview of the approach followed in this study.

2.1. Initial implementation of VPRM

The $VPRM_{INI}$ employs a remote sensing-based scheme to obtain high-resolution flux estimates using the enhanced vegetation index (EVI), and land surface water index (LSWI), derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) reflectance measurements onboard NASA's Terra and Aqua satellites. The vegetation indices from MODIS reported an accuracy within ±0.025 for high-quality observations to retrieve the top of the canopy data (https://modis-land.gsfc.nasa.gov/ValStatus.php?ProductID=MOD13). The VPRM uses annually varying vegetation classification from synergetic land cover product (SYNMAP; Jung *et al* 2006) available from https://databasin.org/datasets/112a942ec4294e5284e63d5e6bf14b29/. According to vegetation classification based on SYNMAP, we considered seven major biomes over the Indian region, which include Cropland (68.2%), Shrubland (16.7%), Evergreen forest (4.8%), Deciduous forest (4.4%), Mixed forest (3.7%), Grassland (1.73%), and Savanna (0.47%) with their areal coverage given in percentage. NEE is calculated based on GPP (light-dependent process) and R_{eco} (light-independent process). We followed the sign convention for NEE in which negative values indicate CO_2 uptake (land sink) and positive values represent CO_2 release into the atmosphere (land source).

$$NEE_{VPRM,INI} = -GPP_{VPRM,INI} + R_{eco,VPRM,INI} \quad (1)$$

The calculation of $GPP_{VPRM,INI}$ and $R_{eco,VPRM,INI}$ is explained in detail in the companion article Ravi *et al* (2024). Also, please refer to the figure 1 of Ravi *et al* (2024)

Model parameters are usually calibrated using site-level EC measurements across different ecosystem types, minimizing the least squares between modeled fluxes and EC observations. This optimization procedure, utilizing discrete tower locations representing major vegetation classes, aims to improve model performance for the region of interest (Luus and Lin 2015, Dayalu *et al* 2017). Due to insufficient observational EC measurements to calibrate the Indian region, we used the standard model parameters, initially optimized against flux measurements representing tropical biomes over the Amazonian region (Botía *et al* 2022) (table S1). However, these parameters may not fully represent subtropical Indian biomes, potentially resulting in reduced model performance compared to VPRM simulations in regions like Europe or North America, where sufficient EC observations are available for calibration. Consequently, we also utilized the other sets of model parameters and examined their impacts on the model performance over the Indian region, which is detailed in the companion article Ravi *et al* (2024).

2.2. Ecosystem uptake refinements using SIF

The performance of the $VPRM_{INI}$ model depends on the density of long-term ground-based observations (e.g. EC) used for calibration. Thus, the unavailability of EC measurements in India hinders the model's reliability. To overcome this limitation, we utilized satellite products based on OCO-2 and TROPOMI, adding more observational constraints to the model, $VPRM_{INI}$, for improving GPP distribution. Table S2 provides an overview of the various datasets used in this study.

We used two global SIF products in this study. Daily SIF retrievals have been available from TROPOMI (hereafter referred to as TROPOSIF, http://ftp.sron.nl/open-access-data-2/TROPOMI/tropomi/sif/v2.1/l2b/; Guanter *et al* 2021) since May 2018. Also, we utilized SIF from GOSIF_v2 (hereafter referred to as GOSIF, http://data.globalecology.unh.edu/; Li and Xiao 2019b), having longer data record than TROPOSIF, providing retrievals at spatial and temporal resolutions of 0.05° and 8 d, respectively. TROPOSIF is a newer product with denser data coverage than GOSIF. The spatial discontinuity in the original daily OCO-2

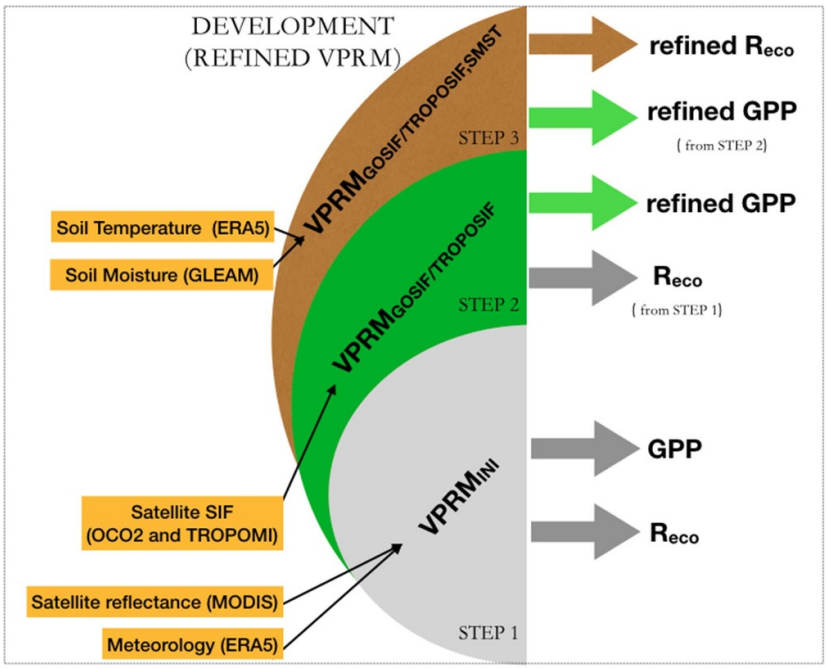

Figure 1. An overview of the methodology followed in the VPRM refinement. CT: carbon tracker, EC: eddy covariance, SIF: solar induced fluorescence, NEE: net ecosystem exchange, GPP: gross primary productivity, R_{eco}: ecosystem respiration.

retrievals is improved in GOSIF using a machine learning (ML) approach based on MERRA-2 meteorological fields, MODIS reflectance, and land cover data, preserving the observed variability of discrete SIF retrievals as explained in Li and Xiao (2019b). Although TROPOSIF is expected to yield better results due to its higher data coverage, we compared these two SIF products and investigated considerable product differences that could lead to variations in GPP estimates over India. A fully calibrated SIF dataset, correcting the errors due to differences in instrument characteristics and retrieval algorithms, is not available owing to the challenges in obtaining reference SIF values representative of a kilometer scale across the world as reported by Parazoo *et al* (2019) and Zhao *et al* (2024). Further, to the best of our knowledge, there are no ground-based SIF measurements in India for conducting local calibration. To tackle this limitation, the separate SIF values in the red and far-red windows have usually been compared for their patterns and consistency. Patterns of red SIF and far-red SIF were found to be consistent, which is also consistent with ground-based measurements Celesti *et al* (2018) and Magney *et al* (2019). Additionally, satellite-based SIF has been used to derive GPP and then calibrated with flux tower GPP across the globe (e.g. Zhang *et al* 2020). Some regional studies in Europe and North America, where extensive EC observation networks are available, have conducted such analyses (e.g. Li and Xiao 2019a, Qiu *et al* 2020, Kunik *et al* 2023). The limited availability of EC flux data in India hinders similar regional analyses over India.

We derived GPP from GOSIF and TROPOSIF, detailed in companion article Ravi *et al* (2024). By including SIF information, $GPP_{VPRM,INI}$ is modified as follows:

$$GPP_{VPRM,SIF}(i,j,t,vg) = \eta_{vg} \times GPP_{VPRM,INI}(i,j,t,vg) + \epsilon_{vg} \quad (2)$$

$GPP_{VPRM,SIF}$ is the modified GPP that includes GOSIF (denoted as $GPP_{VPRM,GOSIF}$) or TROPOSIF (denoted as $GPP_{VPRM,TROPOSIF}$). i, j, and t correspond to latitude, longitude, and time respectively. η_{vg} is the scaling factor corresponding to the specific vegetation class, vg, applied to upscale $GPP_{VPRM,INI}$ to include the information provided by SIF. The η_{vg} values based on TROPOSIF are provided in table S1. ϵ_{vg} represents the error term corresponding to each vegetation class.

2.3. Soil moisture (SM) and temperature in respiration model equation

The soil properties can influence both autotrophic and heterotrophic respiration, particularly in regions with distinct wet and dry seasons (Flexas *et al* 2006, Meir *et al* 2008, Molchanov 2009). A recent study over the Eastern USA and Canada demonstrated improved simulations of R_{eco} by incorporating the effects of changing foliage, water stress, and the non-linear relationship with temperature (Gourdji *et al* 2022). We

utilised the surface SM fields from GLEAM v3 (www.gleam.eu/#datasets; Martens *et al* 2017) model and level 2 (7–28 cm) soil temperature (ST) fields from ERA5 reanalysis product (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview; Hersbach *et al* 2020) (see table S2). These datasets were compared with SM and ST outputs from the high-resolution land data assimilation system (HRLDAS; Chen *et al* 2007) driven by the Noah land surface model (LSM) for the period 2012–2017. Here, we relied on model products due to the limitations of observation data covering the entire domain. The distribution of R_{eco} is re-defined as follows (see the companion article Ravi *et al* (2024) for more details):

$$R_{eco,VPRM,SMST}(i,j,vg) = \tau_{vg} \times ST(i,j,vg) + \nu_{vg} \times SM(i,j,vg) + \kappa_{vg} \times \left(\alpha_{vg} \times T_{air}(i,j,vg) + \beta_{vg}\right) \quad (3)$$

where, τ_{vg}, ν_{vg}, and κ_{vg} represent the vegetation-specific parameters used to improve $R_{eco,VPRM,INI}$ predictions. The observation-based respiration fluxes from FLUXNET-Random forest (https://db.cger.nies.go.jp/DL/10.17595/20200227.001.html.en, Jiye 2020) dataset (hereafter referred to as FLUXNET) are used as a baseline to derive the above parameters. Table S1 provides the details of the vegetation-specific model parameters derived for refining $R_{eco,VPRM,INI}$.

2.4. Other model products for comparison

We used simulated surface CO_2 fluxes from commonly used process-based terrestrial biosphere models for the inter-model comparison and performance assessment. Also, we used the global inversion-based flux estimates consistent with atmospheric mixing ratio observations.

We used simulations from 14 Dynamic Global Vegetation Models within the TRENDYv10 model ensemble for the Indian region (see table S3). All LSMs under TRENDY were driven with common input/forcing data from 1901 to 2020 and followed a common simulation protocol. Model simulations include climate forcing from CRU+CRU-JRA (https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.05/) monthly and 6 h historical forcing for the period 1901–2020, ice core data from 1700 to 2020 and land-use change data from Hyde database for the period 1850–2021. Specifically, this study used TRENDY S3 simulations, which consider the impact of atmospheric CO_2 concentration changes, climate change, and land cover changes on the global terrestrial ecosystem (see https://blogs.exeter.ac.uk/trendy/). Notably, some TRENDY models also incorporate nitrogen cycling; for these models, time-varying nitrogen inputs are provided. Examples include CABLE-POP, CLM5.0, DLEM, ISAM, OCN, and ORCHIDEE-CNP. The TRENDY models used in this study differ in spatial resolution, but each provides fluxes at a monthly temporal resolution.

We used inverse model flux estimates from Carbon Tracker (CT2019B, hereafter referred to as CT) modelling system (https://gml.noaa.gov/ccgg/carbontracker/download.php; Peters *et al* 2007). The biospheric module of CT used prior fluxes from CASA biogeochemical model based on the remotely sensed monthly fraction of Photosynthetically Active Radiation. We used three hourly gridded estimates of optimised biospheric CO_2 fluxes with $1^\circ \times 1^\circ$ horizontal resolution over the Indian domain for the years 2016 to March 2019, available at https://gml.noaa.gov/ccgg/carbontracker/. All these gridded flux estimates used for comparing spatial patterns are aggregated or disaggregated to a common spatial and monthly temporal resolution for comparison (see section 2.6).

2.5. EC flux tower observations for model evaluation

For model evaluation, we used EC observations of terrestrial biosphere CO_2 fluxes from a flux tower located at Betul in Central India (Jha *et al* 2013). Betul tower (commissioned in November 2011) is situated 507 m above mean sea level inside the mixed Deciduous forest where a tropical climate prevails. Generally, the flux towers have small footprints of approximately 1 km radius around the measurement site. The forest in which the Betul flux tower is located covers an area of 176 ha and has a tree density of 400–500 trees ha^{-1}. The forest is homogeneous and free of anthropogenic impacts within a 1 km radius around the tower. The land cover in the model identifies the area ($\sim$100 km^2) around this site as majorly dominated by Deciduous Forest (38.2%), and the remaining biomes in the region are unmanaged homogeneous Crops, Shrubs, and Savanna. The land-cover details are further examined for a zoomed area ($\sim$25 km^2) around the site using higher-resolution Sentinel-2 data (10 m, Brown *et al* 2022), which also supports that the region is mainly surrounded by thick forest-type vegetation (38%) with a significant mixture of unmanaged homogeneous lands, rangeland (see figure S1). These results from SYNMAP and Sentinel-2 thus ensure less impact of land-cover error on our VPRM simulations due to classification mismatches. However, it should be noted that the representation error (due to scale mismatch) is inherent in the simulations, which adds to the model-observation bias, as mentioned earlier.

Uncertainties in the model-observation mismatch due to this spatial scale difference are expected when observations within a 1 km radius are compared with the data from larger areas, such as the 100 km^2 scale of

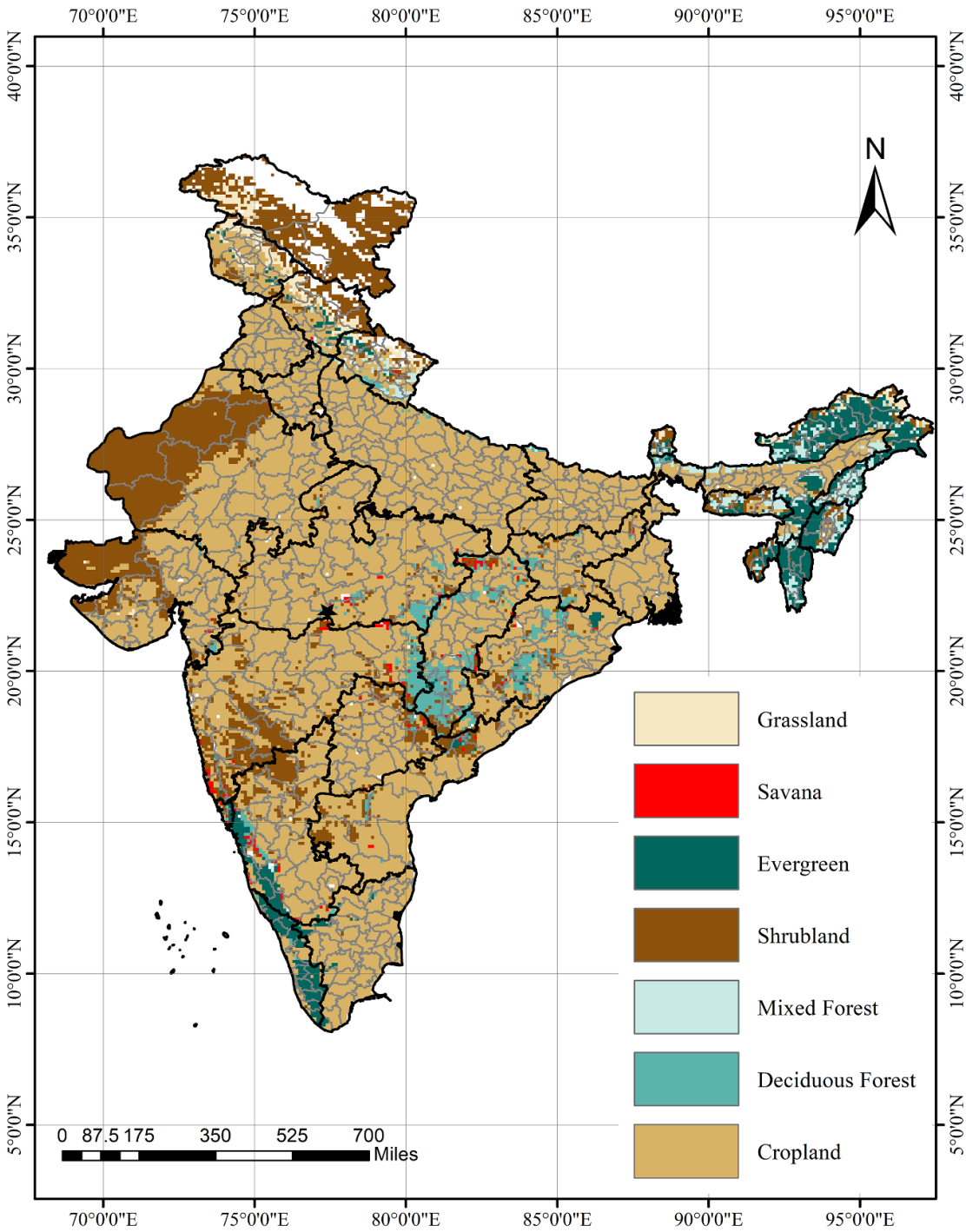

Figure 2. An overview of the major vegetation classes for the study region. The solid black star denotes the EC observation site at Betul.

VPRM. Further descriptions of the site and details of the instrumentation from Betul can be found in published studies (Jha *et al* 2013, Rodda *et al* 2021). Table S4 provides an overview of the characteristics of the flux tower site, and figure 2 shows the location map of the flux tower. The station provides NEE, GPP, and R_{eco} fluxes, where the GPP and R_{eco} are derived from NEE using flux partitioning algorithms as reported by Rodda *et al* (2021). For Betul station the mean annual NEE is reported to be of -0.52 ± 0.040 kg C m^{-2} yr^{-1}, GPP of 3.35 ± 0.16 kg C m^{-2} yr^{-1}, and R_{eco} of 2.83 ± 0.15 kg C m^{-2} yr^{-1} by Rodda *et al* (2021).

The half-hourly data from Betul was aggregated into hourly, monthly, and annual time scales for this analysis. There exist data gaps for specific years. Model simulations are evaluated against observations by taking simulations from the grid cell nearest to the flux tower location (which is 21.89° N, 77.40° E). While model comparison with observations provides valuable insights into model performance, the interpretation must be done cautiously due to scale mismatches between the model and the observations. The

model-observation mismatches due to this spatial scale difference are expected in the present study when observations within a 1 km radius are compared with those from larger areas, such as the 10 km × 10 km scale of VPRM. These spatial scale mismatches in the representativeness of the observations may explain a fraction of the model-observation bias we report in the present study. These representation errors are inherent in simulations but can be reduced when increasing the spatial resolution of the model with highly resolved input data and adequate representation of processes. For example, a recent study has shown the improved performance of VPRM in European cropland flux sites when integrated with high-resolution Sentinel-2 indices with EC measurements (Bazzi *et al* 2024). The future availability of more flux observations representing diverse biomes would enable us to perform a rigorous model evaluation at the ecosystem level. The observation-based fluxes from ML products FLUXCOM (www.bgc-jena.mpg.de/geodb/projects/DataDnld.php, Jung *et al* 2020) and FLUXNET were also used for model comparison.

2.6. Spatial and biome-specific pattern analysis

Here, we used flux simulations generated by refined VPRM to examine spatial gradients and seasonal variations of biospheric fluxes. Since some ecosystems can be more biologically productive than others, we aggregated flux patterns separately for each vegetation class based on SYNMAP land cover types to estimate the productivity of each ecosystem in capturing atmospheric CO_2. Refined VPRM fluxes at a monthly time scale are employed here.

3. Results

As explained in section 2.2, we utilised satellite retrievals of SIF from OCO-2 (GOSIF) and TROPOMI (TROPOSIF) to improve $GPP_{VPRM,INI}$. The spatial and temporal analysis of SIF across Indian biomes and the SIF-GPP relationship derived for each vegetation class is provided in the companion article Ravi *et al* (2024).

3.1. Model evaluation with EC flux observations

Figure 3 compares EC observation from Betul with model simulations. Inter-annual variations exist in GPP, R_{eco}, and NEE fluxes over Betul (figure 3), with significant data gaps in 2014 and 2017. While the observed NEE showed positive values (carbon release to the atmosphere) during summer (March–June), the ecosystem uptake was observed (negative NEE values) for the rest of the year (July–February). $VPRM_{INI}$ shows better agreement in predicting the seasonality in observed monthly averaged NEE fluxes (coefficient of determination (R^2) = 0.59) than CT (R^2 = 0.24) and TRENDY (R^2 = 0.45), but with a significant underestimation of NEE fluxes at a monthly scale (see table S5). Note that we have used the TRENDY model ensemble for comparison, and the variation among TRENDY simulations for NEE (as calculated by the standard deviation from the ensemble mean over the seven years) ranges from −3.50 to 2.47 μmol m^{-2} s^{-1} over Betul. Similar to NEE, the model predicted the monthly mean variations in GPP reasonably well (coefficient of determination (R^2) = 0.71), but with considerable bias (mean bias error (MBE) = −6.7 μmol m^{-2} s^{-1}, root mean square error (RMSE) = 8.3 μmol m^{-2} s^{-1}). Previous studies have shown the underestimation of GPP when MODIS-derived products are used for GPP estimation (e.g. Zhang *et al* 2008). Similar underestimation is also observed when MODIS derived GPP product (MODIS-GPP, see table S2) compared with EC data from Betul (R^2 = 0.58, RMSE = 8.75 μmol m^{-2} s^{-1}, MBE = −6.96 μmol m^{-2} s^{-1}, from 2012 to 2015). Zhang *et al* (2012) and Turner *et al* (2006) have reported that the MODIS-derived GPP shows underestimation for high productive vegetation and overestimation for low productive areas. This bias is particularly pronounced in vegetation types with significant seasonal variability, such as croplands. The MODIS GPP algorithm, which employs a light-use efficiency (LUE) approach, does not account for several factors that influence photosynthesis or for the coupling between water and carbon cycles. This can lead to an inadequate representation of seasonal variations and inconsistencies in the results (Zheng *et al* 2019). Additionally, the MODIS algorithm's use of a constant LUE per vegetation type tends to result in overestimation in low LAI conditions due to shaded leaves and underestimation at high LAI levels (Zhang *et al* 2012, Liu *et al* 2014).

The SIF-based GPP products (GPP_{GOSIF} and $GPP_{TROPOSIF}$ with R^2 = 0.80) are closer than $GPP_{VPRM,INI}$ to the observed GPP in terms of magnitude. But the observed patterns in GPP are better captured by $VPRM_{INI}$ (R^2 = 0.83) (figure S2) than the SIF-based GPP products. This shows the potential of the VPRM model to predict the observed variations in GPP, which leads us to calibrate VPRM model parameters using SIF-based GPP rather than directly using them in our GPP estimations (detailed in the article Ravi *et al* 2024). The $GPP_{VPRM,INI}$ modified based on GOSIF ($GPP_{VPRM,GOSIF}$), and the $GPP_{VPRM,INI}$ modified based on TROPOSIF ($GPP_{VPRM,TROPOSIF}$) are evaluated with EC observations, and the inter-comparison with $GPP_{VPRM,INI}$ shows remarkable improvement in the model performance for GPP with a significant reduction in RMSE and MBE values (see figure 3(a) and table S5).

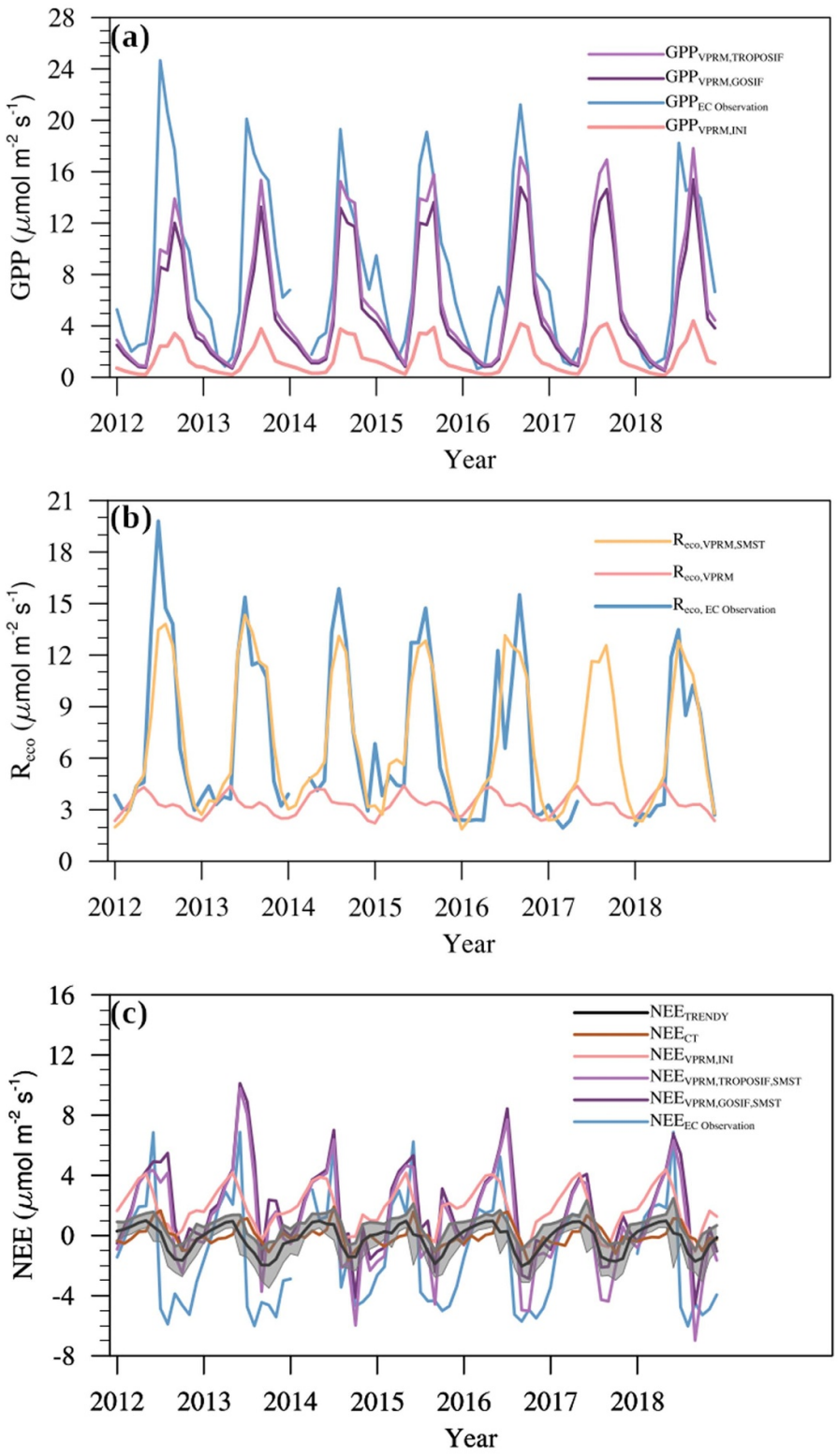

Figure 3. Comparison of monthly averaged EC observations with (a) GPP, (b) R_{eco}, and (c) NEE simulations over Betul for the period 2012–2018. The mean (black line) and ±1 standard deviation (grey shade) of 14 models from the TRENDY model inter-comparison project are provided.

The $VPRM_{INI}$ model fails to capture the seasonality in respiratory fluxes ($R^2 = 0.02$), with a significant underestimation of R_{eco} by -3.5 μmol m^{-2} s^{-1} (RMSE = 5.7 μmol m^{-2} s^{-1}) (see figure 3(b) and table S5). To improve the model performance, we utilised observation-based datasets of ST and SM in addition to the air temperature used for $R_{eco,VPRM,INI}$ simulations. The $R_{eco,VPRM,INI}$ modified based on various datasets (e.g. HRLDAS ST/SM, ERA5 ST, and GLEAM SM) provided similar results. Here, we present the analysis using

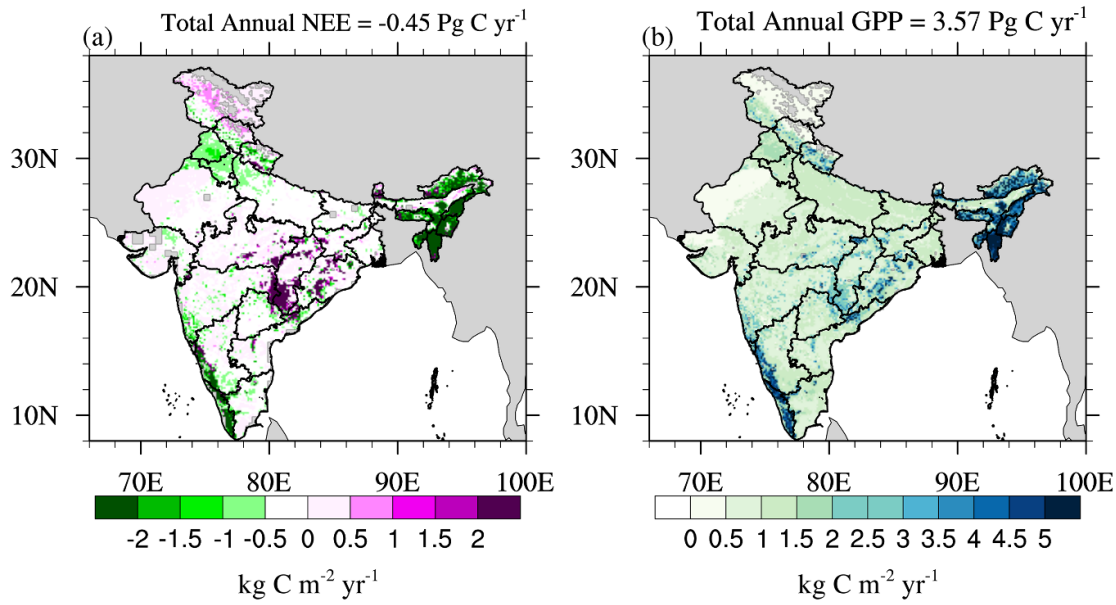

Figure 4. Spatial patterns in annual (a) NEE and (b) GPP fluxes simulated by $VPRM_{REFD}$ over the Indian region averaged for the years from 2012 to 2020.

ERA5 ST and GLEAM SM, considering the large temporal coverage of the data. For the period from 2012 to 2018, the VPRM respiration modified using SM ($R_{eco,VPRM,SM}$) shows much improvement in model prediction than when ST alone is used (detailed in article Ravi *et al* 2024). Using both SM and ST ($R_{eco,VPRM,SMST}$), the model-observation bias reduced considerably, with RMSE reducing from 5.7 μmol m^{-2} s^{-1} to 1.9 μmol m^{-2} s^{-1} and MBE reducing from −3.5 μmol m^{-2} s^{-1} to −0.01 μmol m^{-2} s^{-1}. The statistics from 2018 provided in table S5 also support the above findings.

The VPRM NEE estimated based on $GPP_{VPRM,GOSIF}$ and $R_{eco,VPRM,SMST}$ (hereafter referred to as $NEE_{VPRM,GOSIF,SMST}$) and based on $GPP_{VPRM,TROPOSIF}$ and $R_{eco,VPRM,SMST}$ (hereafter referred to as $NEE_{VPRM,TROPOSIF,SMST}$) are evaluated with EC observations (figure 3(c)). The modified models showed improvement over $NEE_{VPRM,INI}$ in capturing the observed seasonal pattern with a reduction in errors during the period from 2012 to 2018 (RMSE: $NEE_{VPRM,GOSIF,SMST}$ = 4.4 μmol m^{-2} s^{-1}, $NEE_{VPRM,TROPOSIF,SMST}$ = 3.8 μmol m^{-2} $s^{-1}$and MBE: $NEE_{VPRM,GOSIF,SMST}$ = 3.2 μmol m^{-2} s^{-1}, $NEE_{VPRM,TROPOSIF,SMST}$ = 2.4 μmol m^{-2} s^{-1}) (see table S5). The observed seasonal anomalies in NEE ranges from −4.9 to 8 μmol m^{-2} s^{-1} with a standard deviation of 3.6 μmol m^{-2} s^{-1}. These variations are well captured by our model with a mean bias of 1.6 μmol m^{-2} s^{-1}. The modifications made to GPP and R_{eco} fluxes in VPRM improved the model's ability to capture NEE fluxes over Betul. Since $VPRM_{TROPOSIF,SMST}$ (providing $NEE_{VPRM,TROPOSIF,SMST}$, $GPP_{VPRM,TROPOSIF}$, and $R_{eco,VPRM,TROPOSIF,SMST}$ distribution) is found to be closer to the observation among other modified VPRM models, the rest of the analysis uses the simulations from $VPRM_{TROPOSIF,SMST}$ (hereafter referred to as $VPRM_{REFD}$).

With the refined model $VPRM_{REFD}$, the NEE estimates for Betul range from -1.46 ± 0.71 kg C m^{-2} yr^{-1} to 2.02 ± 0.79 kg C m^{-2} yr^{-1}, aligning closely with observed values of -1.91 ± 0.39 kg C m^{-2} yr^{-1} to 2.04 ± 0.81 kg C m^{-2} yr^{-1} (Rodda *et al* 2021). At Kaziranga National Park, a deciduous broadleaf forest in northeast India, our estimates suggest the region as a net carbon source within the range of 0.65 ± 0.06 kg C m^{-2} yr^{-1} to 1.11 ± 0.13 kg C m^{-2} yr^{-1} annually (see table S9), compared to observed values ranging from −0.029 kg C m^{-2} yr^{-1} to 0.207 kg C m^{-2} yr^{-1} (moderate carbon source) reported by Burman *et al* (2021). At the Kozi Katarmal evergreen forest site, $VPRM_{REFD}$ estimates range from -1.46 ± 0.18 kg C m^{-2} yr^{-1} to 0.42 ± 0.16 kg C m^{-2} yr^{-1} reporting the region predominantly acts as a net carbon sink (see table S9) while EC measurements also indicate it as carbon sink, ranging from −0.64 kg C m^{-2} yr^{-1} to −0.9 kg C m^{-2} yr^{-1} (Burman *et al* 2021).

3.2. Flux spatial patterns

We find strong spatial variations in the annual NEE and GPP distribution by $VPRM_{REFD}$ over the Indian region (see figure 4), with distinct zonal and meridional variations. Annually, most parts of the country remained as carbon sinks, particularly the southwest and northeast, which are dominated by Evergreen and Mixed forests. They also showed higher GPP values, indicating high productivity. The eastern part of central

India dominated by Deciduous ecosystems also showed high GPP values, but respiration surpassed productivity, making the region a carbon source annually. A major part of the country shows moderate GPP values (∼0.05 kg C m^{-2} yr^{-1}–1.5 kg C m^{-2} yr^{-1}), while a large area is covered by Croplands. Ecosystem productivity is minimal in the northern and northwestern parts of the country under Shrubland vegetation.

During our study period (2012–2020), the Indian terrestrial biosphere acted as a net carbon sink annually. The NEE(GPP) value increased from −0.38 Pg C yr^{-1}(3.39 Pg C yr^{-1}) in 2012 to −0.53 Pg C yr^{-1}(3.88 Pg C yr^{-1}) in 2020 (table S6). The multi-year mean annual NEE and GPP estimates from $VPRM_{REFD}$ over the Indian region is -0.45 ± 0.05 Pg C yr^{-1} and 3.57 ± 0.14 Pg C yr^{-1}, respectively (figure 4). The model inter-comparison shows that the total NEE fluxes are lower in CT and TRENDY compared to those of $VPRM_{REFD}$ (μ($VPRM_{REFD}$-TRENDY) = −0.34 Pg C yr^{-1}; μ($VPRM_{REFD}$-CT) = −0.25 Pg C yr^{-1} in which μ represents sample mean of differences, statistics presented for the year 2018). An ensemble average using 14 TRENDY models was used for the analysis. Our annual NEE estimates show higher uptake than the previously published studies in which process-based and LUE models were used (Nayak *et al* 2015, Cervarich *et al* 2016, Rao *et al* 2019). Based on the CASA model, Nayak *et al* (2015) estimated a NEE value of −0.01 Pg C yr^{-1} for a 26 year period from 1981 to 2006, showing ecosystem transition from a carbon source in the 1980s to a carbon sink in the subsequent decades. Using the TRENDY model ensemble, Cervarich *et al* (2016) and Rao *et al* (2019) estimated annual NEE values of −0.2 Pg C yr^{-1} and −0.14 Pg C yr^{-1}, respectively for India.

The spatial patterns for monthly averaged NEE and GPP are presented in figure 5. Carbon uptake shows the highest value during the summer monsoon season (August–October) and the lowest value during the dry and hot months from April to June. The Indo-Gangetic plain shows high carbon sink capacity during the winter (January–February) and summer monsoon months (August–September). The southeastern part of the country shows an increase in productivity during the northeast winter monsoon (October–December), which is reflected in the GPP and NEE values. Most parts of the country remained carbon neutral from April to May except the Deciduous forest of central India and Evergreen Forests on the southwestern and northeastern sides.

3.3. Model comparison with observation-based products

The FLUXCOM and FLUXNET are two widely accepted observational data-driven ML products in representing the vegetation carbon dynamics (Friend *et al* 2007, Melaas *et al* 2013, Jung *et al* 2020, van Dijke *et al* 2020, He *et al* 2022, Jiao *et al* 2024). Therefore, we compare $VPRM_{INI}$ and $VPRM_{REFD}$ simulations with those from FLUXCOM and FLUXNET (figure S3) to examine the overall agreement between our GPP products and these global products. Since we used FLUXNET to scale the R_{eco} fluxes, we restrict our analysis to GPP only here. In general, we find that the $VPRM_{REFD}$ is close to FLUXNET and FLUXCOM compared to $VPRM_{INI}$ in terms of capturing the seasonality and magnitude of fluxes. The cropland occupies around 68.2% of the total land area of India; hence, we also specifically analysed fluxes from Cropland where $VPRM_{REFD}$ showed a significant reduction in RMSE and MBE compared to $VPRM_{INI}$ (table S7).

Although the reduced inter-model differences between $VPRM_{REFD}$ and observational data-driven products are encouraging, it should be noted that neither FLUXCOM nor FLUXNET has incorporated the EC observations from Indian biomes. In this case, we have chosen EC measurements to evaluate how well our models represent Indian biospheric fluxes. To do this, we compared FLUXCOM and FLUXNET data with long-term records of EC observations from Betul, assessing the extent to which they captured the observed variability (figure S3). FLUXNET ($R^2 = 0.80$) captured the pattern in observed fluxes better than FLUXCOM ($R^2 = 0.56$), but the bias exhibited by both models is in a similar range. Our $VPRM_{REFD}$ outperformed FLUXCOM and FLUXNET in capturing the observed flux distribution.

Additionally, observation-based ML products, such as FLUXNET and GOSIF_GPP, are used to compare NEE and GPP fluxes against $VPRM_{INI}$ and $VPRM_{REFD}$ at selected pseudo-EC flux tower sites in India (see figures S4 and S5).

3.4. Derived ecosystem productivity and exchanges across different biomes

We have analysed the derived GPP and NEE across seven vegetation classes used in VPRM for 2020 (table S8). On an annual scale, the Mixed forest vegetation shows the highest (GPP = 6.35 kg C m^{-2} yr^{-1}) productivity, followed by the Evergreen forest, Deciduous forest, and Savanna biomes (GPP = 5.51 kg C m^{-2} yr^{-1}, 4.63 kg C m^{-2} yr^{-1}, 4.60 kg C m^{-2} yr^{-1}, respectively) during 2020. As expected, lower productivity rates are found for Shrubland (1.74 kg C m^{-2} yr^{-1}) and Cropland (1.43 kg C m^{-2} yr^{-1}). Even though Cropland covers a large fraction of Indian land mass, its GPP per unit area is lower compared to that of Deciduous Forests, Evergreen Forests, and Mixed Forests, despite the smaller total land area covered by these vegetation types. The annual productivity for Grasslands is the lowest with a GPP value of 0.66 kg C m^{-2} yr^{-1}. Even though Deciduous vegetation shows high GPP values, R_{eco} dominates GPP across this biome,

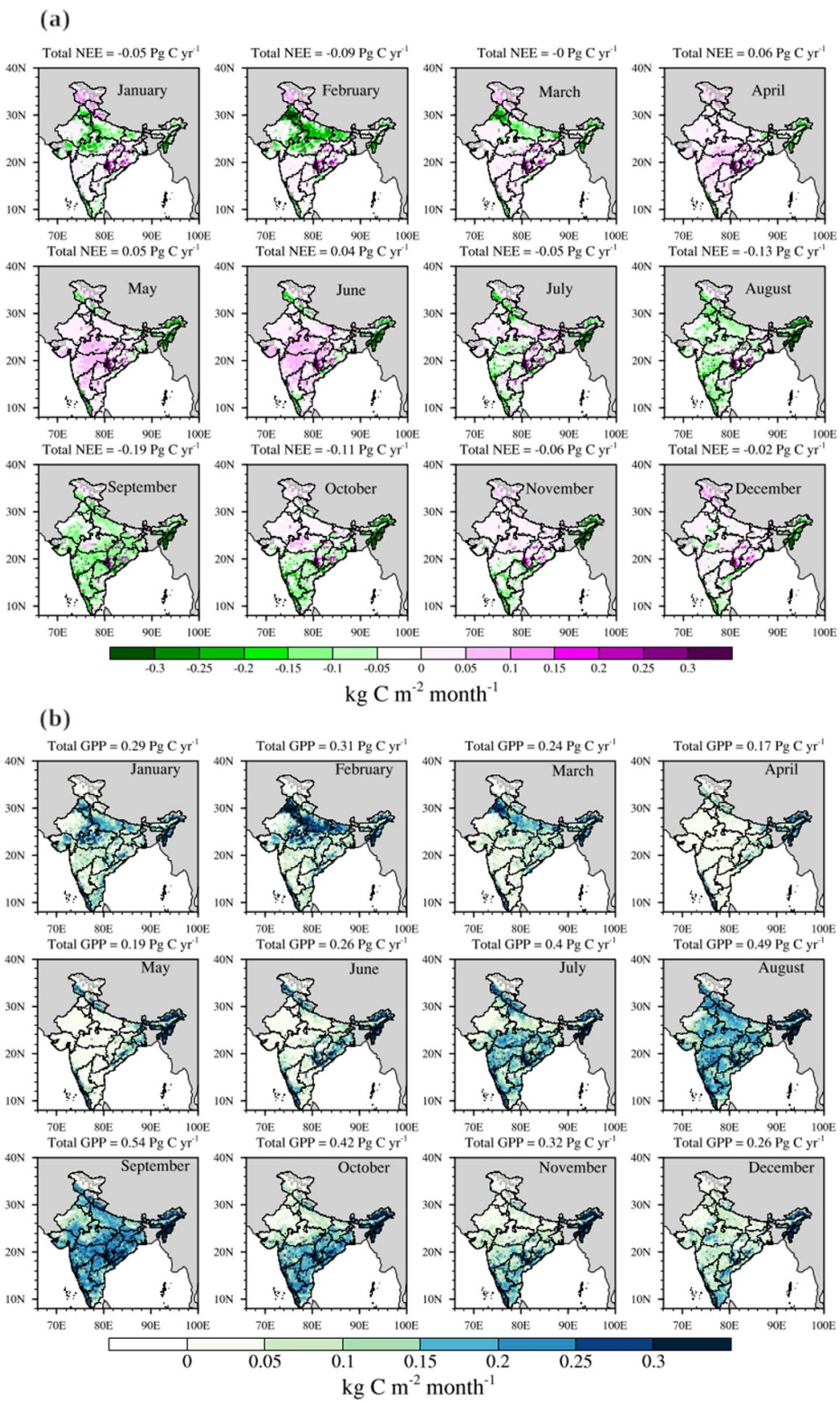

Figure 5. Spatial pattern in the monthly averaged fluxes of (a) NEE and (b) GPP from $VPRM_{REFD}$ for the year 2020.

leaving it as a carbon source or carbon neutral on an annual scale (Burman *et al* 2021, Sarma *et al* 2022). The highest productivity by forests over Grassland is also seen in other parts of the globe (Yu *et al* 2013).

The contribution of each vegetation to the national GPP also depends on the area covered by each vegetation. As a result, to the national GPP value of 3.88 Pg C yr^{-1}, for the year 2020, Cropland is the major contributor (49.6%), followed by Evergreen forest (14.9%), Mixed forest (12.2%), Shrubland (12.0%), Deciduous forest (9.7%), Savanna (1.1%) and Grassland (0.5%). On an annual scale, Mixed and Evergreen

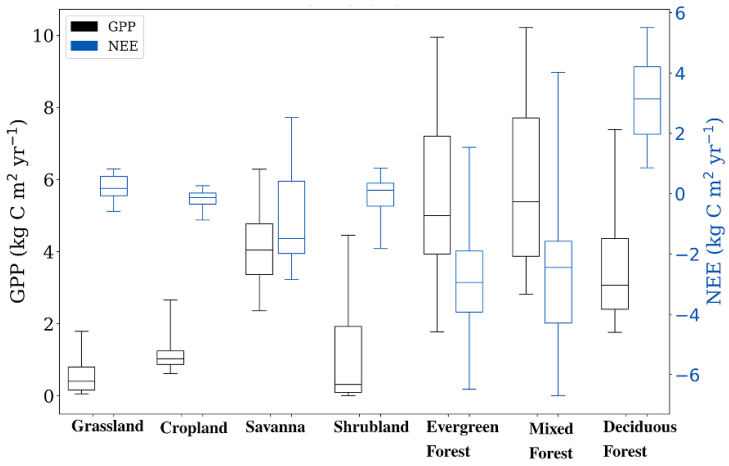

Figure 6. The annual GPP (black box) and NEE (blue box) per m^2 of each biome in India simulated by the $VPRM_{REFD}$ averaged over the period 2012–2020. The upper and lower limit of the box shows the 5th and 95th percentile of the data and the center line shows the median. All the values that are 1.5 times higher than the 5th and 95th percentile are considered outliers and are removed from the graph.

forest vegetation show large GPP variability, while Cropland and Grassland exhibit lower GPP variability (figure 6(a)).

The Evergreen and Mixed forest vegetation have the highest carbon fixation sink capacity, showing high negative NEE values close to ~ -2.5 kg C m^{-2} yr^{-1}, followed by Savanna with an annual NEE value of ~ -1.31 kg C m^{-2} yr^{-1} (table S8). A moderate net carbon fixation efficiency (NEE of ~ -0.39 and ~ -0.28 kg C m^{-2} yr^{-1}) is shown by Shrubland and Cropland vegetations, respectively. The lowest net carbon fixation efficiency is found for Deciduous vegetation, followed by Grasslands. On an annual scale, Mixed forest and Savanna vegetation classes show large NEE variability while lowest by the Cropland and Grassland (figure 6(b)).

Based on the statistical analysis for the year 2020, the Cropland and Evergreen forests are the largest contributor (also considering the area of coverage), in terms of CO_2 sink capacity, to the national NEE budget ($\sim$63.5%) followed by Mixed forest ($\sim$21.2%), Savanna ($\sim$13.5%), Shrubland ($\sim$1.4%), Grassland ($\sim$0.3%), and Deciduous forest ($\sim$0.1%).

3.5. Seasonal and diurnal cycles across different biomes

Figures 7 and 8 shows the seasonal variations in $VPRM_{REFD}$ simulated NEE and GPP fluxes across different biomes averaged from 2012 to 2020. The seasonality in NEE and GPP varied across the vegetation showing a transition from dry and cooler winters to wet and hot summers. Most vegetation showed higher carbon uptake during the monsoon period (negative NEE and high GPP) while they remained as carbon source (positive NEE and low GPP) during the summer months (March–May). The ecosystem GPP for croplands showed a semi-annual cycle with a primary productivity peak during the winter months (December–January) and a secondary peak during the monsoon season (August–September). Savanna vegetation showed strong seasonality in flux pattern (NEE: 12.6 μmol m^{-2} s^{-1}; GPP: 17.1 μmol m^{-2} s^{-1}), while Grassland displayed the least (NEE: 2.2 μmol m^{-2} s^{-1}; GPP: 3.4 μmol m^{-2} s^{-1}). The interannual variability in the seasonal cycle (see the standard deviation curve in figures 7 and 8) is greater for forests than for other biomes. This variation may be attributed to the differing responses of forest carbon fluxes to climate extremes, land use changes, natural disturbances, forest product harvesting, and vegetation growth (Li *et al* 2021).

Figure 9 shows the diurnal variations in $VPRM_{REFD}$ simulated monthly GPP fluxes for different vegetation classes during 2020. The diurnal variability of GPP varies with the season. The peak uptake time is found to vary with vegetation. Larger productivity is found during noon hours (10:00–14:00 local time). Strong daytime variability (peak uptake during noon hours and weak uptake during afternoon hours (15:00–20:00 local time)) is found for Grassland, followed by Shrubland, Savanna, and Cropland vegetation. From the standard deviation curves, vegetation such as Grasslands, Cropland, Savanna, and Shrublands show large monthly variability in the diurnal cycle, while forests have the least variability.

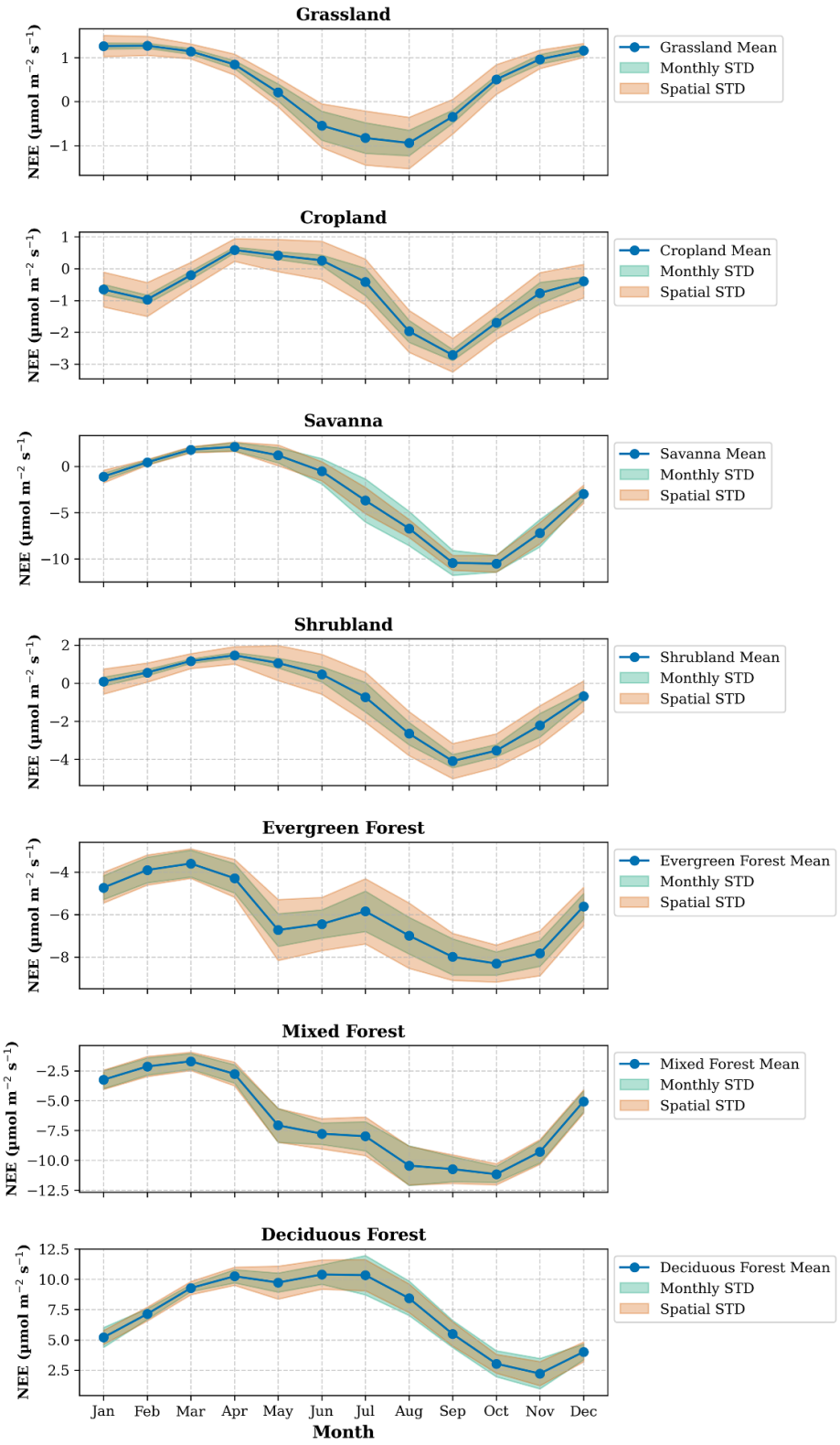

Figure 7. Temporal variations in monthly averaged NEE fluxes from $VPRM_{REFD}$ averaged from 2012 to 2020. The solid blue line indicates the multi-year mean and the shaded region indicates ± 1 standard deviation of temporal and spatial variability.

4. Discussions

Betul is a tropical deciduous forest, and hence, the strong seasonality observed in the fluxes is likely linked to changes in plant physiology. GPP fluxes from $VPRM_{INI}$ were underestimated, which can be attributed to the use of MODIS reflectance products. While these MODIS products capture vegetation dynamics more effectively than SIF, they fall short in accurately representing the magnitude of fluxes when compared to

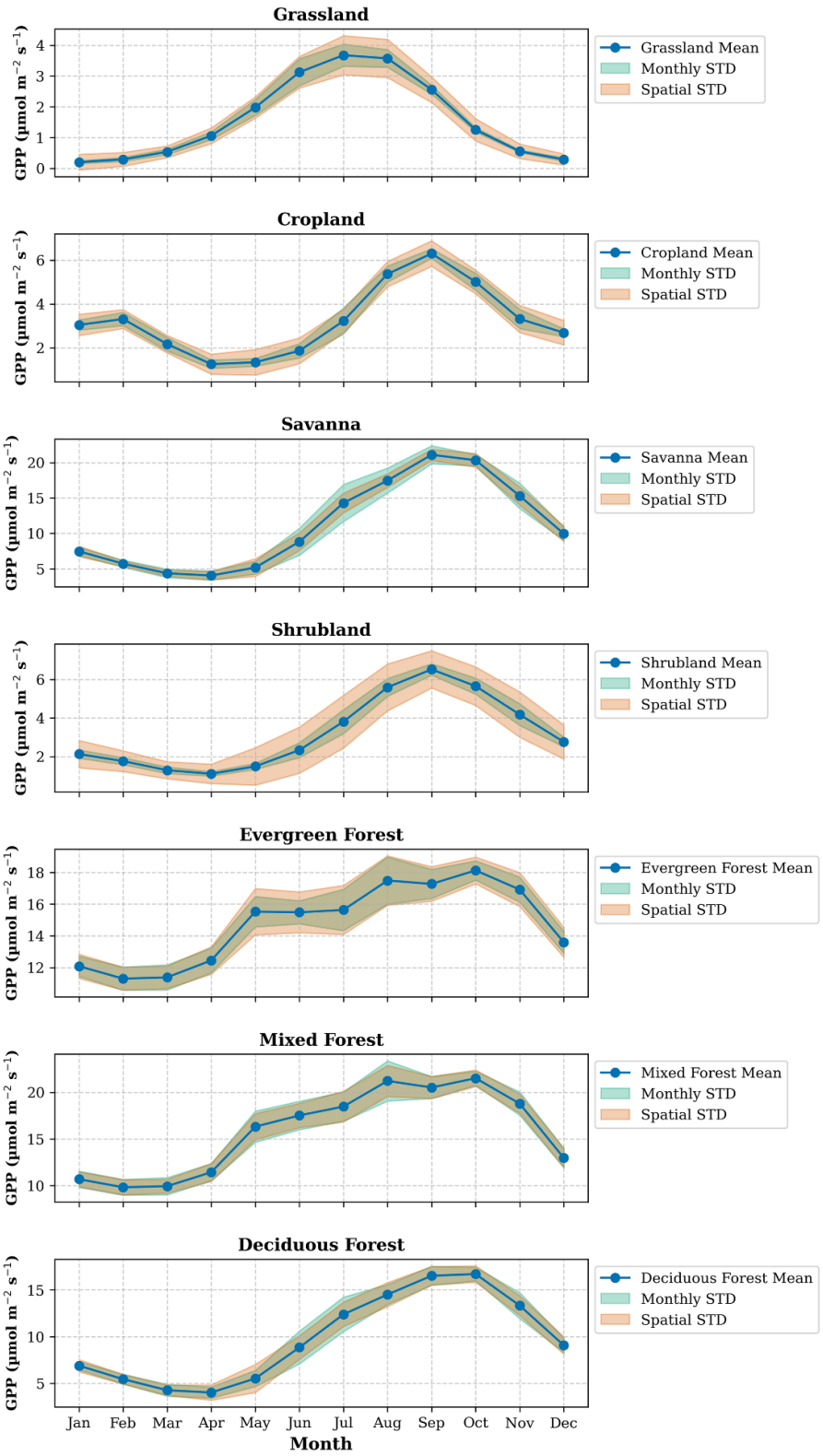

Figure 8. Temporal variations in monthly averaged GPP fluxes from $VPRM_{REFD}$ averaged from 2012 to 2020. The solid blue line indicates the multi-year mean, and the shaded region indicates ±1 standard deviation of temporal and spatial variability.

observations (see figure S2). Nonetheless, $VPRM_{INI}$ demonstrates a better ability to capture seasonal patterns in NEE and GPP compared to other biospheric models, such as the inversion product CT and the ensemble of process-based models TRENDY. It is important to note that these global models (CT and TRENDY) operate at a coarse spatial resolution, and when compared with site-level data (maximum representativeness of only 1 km^2), the scale mismatch between models and observations may significantly contribute to the

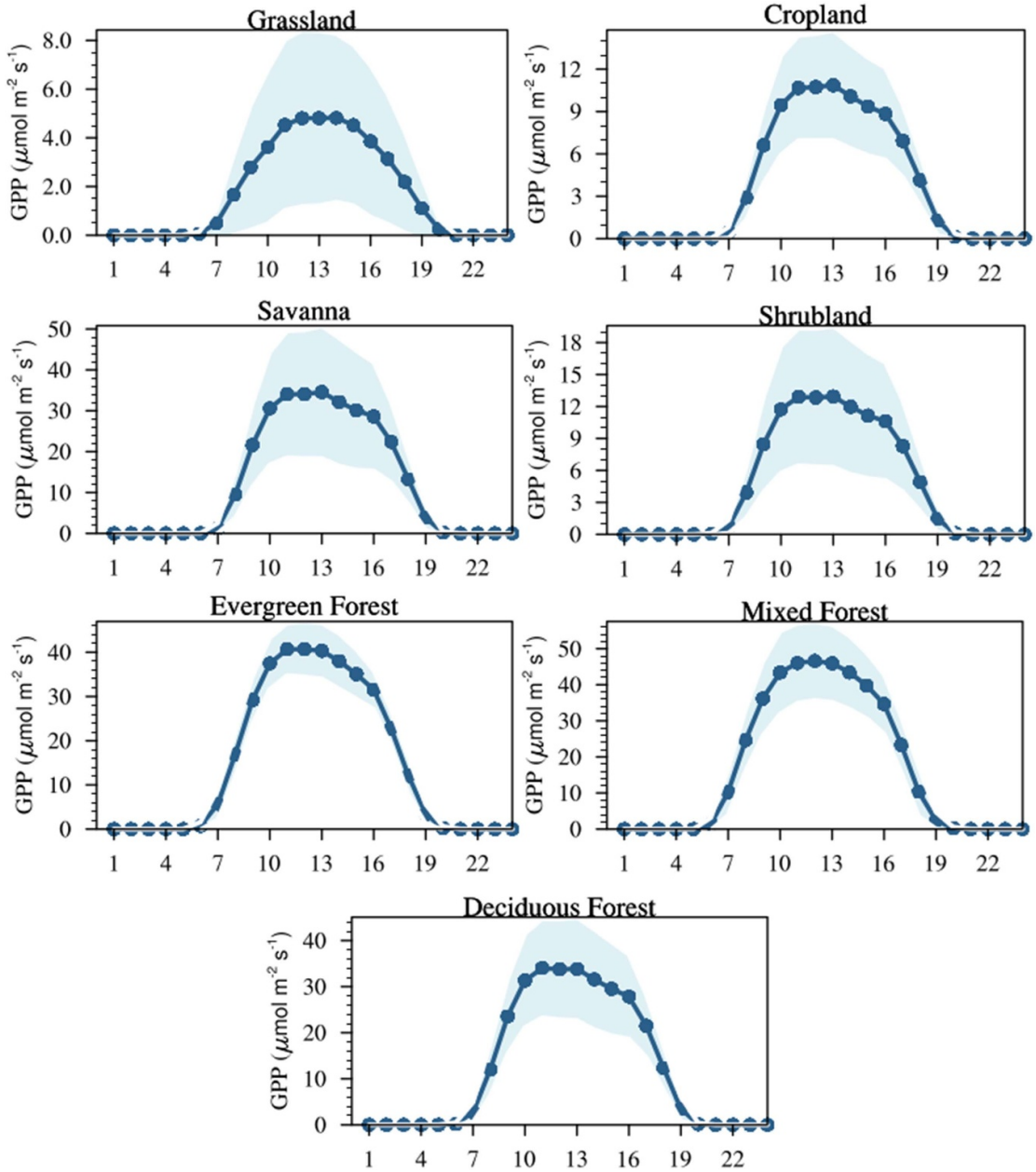

Figure 9. Diurnal variations in $VPRM_{REFD}$ GPP fluxes during 2020. The solid blue line indicates the annual mean for different vegetation classes and the shaded region indicates ±1 standard deviation of the monthly variations.

models' reduced performance. This highlights a potential limitation when applying these models in regional scale applications. Similar to TRENDY models, the $VPRM_{REFD}$ showed reduced capability when performed at a coarser resolution of 0.5° × 0.5° (RMSE = 7.53 μmol m^{-2} s^{-1}, and MBE = 5.74 μmol m^{-2} s^{-1}). The model's high spatial resolution is thus crucial for accurately resolving flux variability.

A significant drawback of $VPRM_{INI}$ is the lack of calibration of model parameters against EC flux measurements in India, as conventionally practiced in regions such as Europe or North America. The unavailability of long-term flux observations across different biomes over the Indian sub-continent is thus a limiting factor for choosing sufficient calibration sites for the $VPRM_{INI}$ parameter tuning, affecting the model performance.

Incorporating SIF in VPRM has noticeably improved the ability of the model to capture the observed seasonal variability of GPP (see figure 3). Also, as a direct proxy for photosynthesis, SIF is expected to provide improved estimates than conventional vegetation indices (Zhang *et al* 2016) (e.g. EVI, LSWI) used in $VPRM_{INI}$ GPP estimation. The observed seasonal anomalies in GPP (variability after subtracting the decadal mean), associated with ecosystem stress and phenology, range from −7.0 to 17.0 μmol m^{-2} s^{-1} with a standard deviation of 6.3 μmol m^{-2} s^{-1}. These variations are well captured by our model with a mean bias

of -1.8 μmol m^{-2} s^{-1}. For instance, The above levels of model improvements confirm the potential of using high-resolution satellite-derived SIF to capture the seasonal cycle of GPP at an ecosystem level. Hence, our results are broadly consistent with the findings by Joiner *et al* (2018), Qiu *et al* (2020)and Wood *et al* (2017) where incorporating SIF data improved GPP estimates.

Incorporating the ST and SM in addition to air temperature for R_{eco} calculation in VPRM improves the model's ability to simulate more realistic values over the Deciduous ecosystem of Betul. The improvement in VPRM R_{eco} while incorporating ST is also reported elsewhere (e.g. Luus and Lin 2015). However, the model simulations can be impacted by the spatial scale mismatch between the VPRM simulations and soil variables used for the model refinements. This needs to be considered while interpreting our results and the model comparisons. While we approximate 85% confidence on our GPP estimations based on input data uncertainties, the confidence can be reduced to 65%–75% (relative error) over Deciduous Forest while comparing them with point measurements (e.g. EC observations) due to other model uncertainties (e.g. inadequate representation of processes and spatial mismatches). Because of these, we approximate less confidence in our R_{eco} estimations (60%–75%). These uncertainties can be minimized by the availability of accurate spatial distributions of soil-related variables at high spatial and temporal resolutions. The modifications made to GPP and R_{eco} fluxes in VPRM improved the model's ability to capture NEE fluxes over Betul.

It should be noted that exclusively data-driven methods using ML tools on observational data (e.g. EC measurements) can be a valuable tool in representing carbon fluxes (Dou *et al* 2018, Joiner *et al* 2018, Zheng *et al* 2020, Lu *et al* 2021, Virkkala *et al* 2021), provided these methods are also subjected to uncertainties and limitations. While these methods are immensely useful in proving gap-filled carbon flux data, some major issues remain, arguably due to heavy dependence on the historical fluxes without advancing the process understanding. Recent studies have pointed out the considerable uncertainties in the inter-annual variability of the carbon uptake estimated by ML methods (Jung *et al* 2020, De Bartolomeis *et al* 2023). Nevertheless, given the availability of an adequate amount of training data, future research using ML methods can be planned to generate complementary datasets. These datasets can thus be used to correct the potential biases in the physics-based models, following the recent work of Upton *et al* (2024).

The analysis of flux spatial patterns revealed that the country exhibited notable spatial variations in NEE and GPP fluxes. These variations are primarily influenced by factors such as the annual mean patterns of temperature, precipitation, and radiation, which play a significant role in shaping the spatial distribution of ecosystem carbon fluxes (Yu *et al* 2013). Despite high GPP values, the deciduous forests of central India acted as a carbon source, as respiration exceeded productivity, leading to a net annual carbon release. The high negative NEE and positive GPP values observed across the Indian region from August to October indicate significant sink capacity, which is linked to increased agricultural crop production driven by monsoonal rainfall. In contrast, winter crop harvesting and unfavorable conditions for plant growth, such as high temperatures, low water availability, and reduced SM content, lead to minimal productivity during the April–May period.

The biome-specific analysis revealed that forests sequester more carbon through photosynthesis than other vegetation groups. Forests have always shown larger efficiency in storing carbon than other vegetation types (McKinley *et al* 2011, Gray and Whittier 2014). However, croplands are the primary contributors to the national carbon budget, largely due to their extensive coverage. In terms of net NEE, Evergreen and Mixed forests exhibit the highest carbon fixation sink capacity, whereas Deciduous forests and Grasslands demonstrate the lowest. Regarding the national NEE, Croplands and Evergreen forests significantly contribute to negative NEE values, while Deciduous forests contribute the least.

The following factors can explain the seasonality observed in India's major vegetation: During the monsoon season, increased water availability stimulates vegetation growth, resulting in GPP surpassing R_{eco}, making the vegetation a carbon sink. In contrast, the hot and dry summer months increase R_{eco} for plant maintenance and hinder vegetation development, leading to a net carbon source (Valsala *et al* 2013). Croplands, however, exhibit a double peak in growing seasons, corresponding to India's two main cropping periods (rabi-winter and kharif-summer), which is also reported by Goroshi *et al* (2014) and Varghese and Behera (2019). The primary GPP peak occurs in February during the winter season, with a secondary peak in September during the summer monsoon. High productivity during the summer monsoon is driven by substantial rainfall across the country, with GPP beginning to increase in June and peaking in September (see figure S6). As the pre-monsoon season (March–May) begins in March, GPP declines, reaching its lowest point in April. Despite average rainfall, low productivity during the pre-monsoon season is likely attributed to severe water stress from elevated temperatures. Local agricultural practices, including irrigation and nitrogen fertilization, also influence crop development and carbon exchange on a local scale. These effects vary regionally based on species type and farming practices. However, on a national scale, Cropland GPP aligns more closely with the seasonality of water availability through precipitation (see figure S6), also

reported by Varghese and Behera (2019). Most Indian biomes show high GPP values in response to rainfall during the monsoon. Furthermore, most Indian vegetation shows high carbon uptake, increasing GPP when direct sunlight is available (Rodda *et al* 2016, Sarma *et al* 2022).

The main limitation of the present study is the lack of EC observations over India, which are essential for model validation and parameter calibration. This reliance on parameters calibrated for Amazonian and European biomes can introduce uncertainty, further compounded by cloud cover during the monsoon season, which hinders satellite data acquisition. While we attempted to evaluate the model using EC observations from Deciduous Forests and compared it with other process-based models, future research should focus on using a more comprehensive set of flux site observations and atmospheric CO_2 measurements. These should be strategically selected to cover India's diverse geography, climate, and vegetation types to test the model's robustness and identify any mismatches in process-based understanding. Additionally, the study assumes a linear relationship between SIF and GPP across all vegetation types and seasons despite some evidence elsewhere suggesting a non-linear relationship influenced by physiological factors. However, there are no ground-based SIF measurements in India that are co-locating with GPP measurements, which hinders the direct derivation of SIF-GPP relationships across biomes under diverse environmental conditions in India. Further, probable errors in satellite RS product extraction of SIF values under all sky conditions can contribute additional uncertainties to our results. The study also notes that vegetation rapidly responds to environmental changes, and the satellite data used—8 to 16 d composites—may miss significant events or processes, especially in agricultural fields, which can change rapidly. The long revisit time of satellites may miss critical plant responses during these transitions.

While our approach integrated SM and ST with air temperature in VPRM R_{eco} calculations to improve the simulations, future work is required to examine the impact of factors such as moisture and heat stress, phenology, biological influences, and biomass disturbances on R_{eco}.

5. Conclusions

This study presents the improved terrestrial flux distribution of CO_2 over India on a $0.1^\circ \times 0.1^\circ$ grid at a temporal resolution of 1 h from 2012 to 2020. We utilise satellite reflectance products and high-resolution meteorological data in a data-driven biospheric model, VPRM to improve the model estimates of terrestrial biosphere CO_2 flux components over India. In particular, we take advantage of satellite missions, such as TROPOMI and OCO-2, which provide SIF retrievals and relate them to ecosystem productivity across different biomes. The derived flux products explain the magnitude and fine-scale variability over the region better than other model estimates.

Our standard VPRM model captures the observed seasonal patterns in NEE ($R^2 = 0.59$) and GPP ($R^2 = 0.71$) well compared to other biospheric models with different model structures, such as the inversion product CT ($R^2 = 0.24$) and the ensemble of process-based models from TRENDY ($R^2 = 0.45$). But with considerable underestimation in magnitudes on a monthly scale. By improving GPP and R_{eco} simulations, the model has improved its ability to capture the observed NEE fluxes with a significant reduction in model biases. Our analysis showed the highest productivity by forests and the lowest by Shurblands. Despite high productivity, Deciduous forests are an annual carbon source due to elevated respiratory fluxes. Since more than 68.2% of the country is covered with Croplands, the agricultural pattern influences the seasonality in GPP and NEE. Overall, we find that the Indian biosphere acts as a sink with an annual NEE ranging from -0.38 Pg C yr^{-1} to -0.53 Pg C yr^{-1} and an annual GPP ranging 3.39 Pg C yr^{-1}–3.88 Pg C yr^{-1} for the years from 2012 to 2020.

Given the considerable difference in flux components among the terrestrial biospheric models, the analyses demonstrated here can guide future model improvements in deriving GPP and R_{eco}. By showing the potential of the VPRM model to predict the observed variations in GPP better than solely SIF-based GPP products, the present study demonstrates the way to calibrate the model parameters in the absence of EC measurements.

Potential improvements to VPRM include (i) further refinement in the R_{eco} accounting for moisture and heat stress and other biomass disturbance and (ii) incorporating flux observations from different ecosystems to enhance the flux representativeness with better empirically derived and biome-specific model parameters. Given the availability of additional EC observations and satellite retrievals, we anticipate further refinements in our current model. While we report fluxes at a 10 km spatial scale in this study, future studies can focus on conducting VPRM simulations at site or ecosystem scales when additional site-specific observations (e.g. EC measurements) and satellite input fields to match such spatial resolution (e.g. 1 km) become accessible. Performing VPRM at a finer resolution with more observations would likely reduce the currently reported model uncertainties and enhance its potential to capture the finer flux variations over India. The increased availability of future flux tower observations will help optimise the model parameters to enhance the

robustness of these simulations. The next step would be to combine atmospheric observations of CO_2 in an inverse modelling framework together with our derived flux maps to better understand the Indian carbon budget. The demonstrated approach is scalable to other regions, especially where ground-based EC observations are the current limitation. Future research can also focus on incorporating high-resolution land-cover maps such as those from Sentinel-2 into VPRM to quantify its impact on improving the representation of flux variations (Ienco *et al* 2019, Brown *et al* 2022, Bazzi *et al* 2024).

Data availability statement

The VPRM simulations will be made available upon request to the corresponding author. The Carbon Tracker (CT2019B) is freely available online at https://gml.noaa.gov/ccgg/carbontracker/CT2019B/. TRENDYv10 datasets used in this study are available upon request to S Sitch. Eddy covariance observation data may be available upon request to NRSC; www.nrsc.gov.in/. The TROPOMI data is available online at http://ftp.sron.nl/open-access-data-2/TROPOMI/tropomi/sif/v2.1/l2b/. GOSIF_v2 datasets used are available freely from http://data.globalecology.unh.edu/. ERA5 data used is freely available at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview. GLEAM v3 data is available freely at www.gleam.eu/#datasets. FLUXNET data is available freely from https://db.cger.nies.go.jp/DL/10.17595/20200227.001.html.en. FLUXCOM data used is freely available from www.bgc-jena.mpg.de/geodb/projects/DataDnld.php.

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.5281/zenodo.10245534.

Acknowledgments

This study has been funded by the Max Planck Society allocated to the Max Planck Partner Group at IISERB. DP acknowledges the support from the Science and Engineering Research Board (SERB) through an Early Career Research Award (Grant No. ECR/2018/001111) for generating some data products used in the study. AR acknowledges the support of IISERB's high-performance cluster system for computations, data analysis, and visualization. AR and VT are grateful to the Ministry of Human Resource Development (MHRD, India) for their PhD scholarships. We thank National Remote Sensing Centre (NRSC), Hyderabad, for providing access to Betul EC flux tower data, and we acknowledge the efforts of scientists and technicians from the Forestry and Ecology Group at NRSC Hyderabad for the EC data acquisition. The authors thank editor and reviewers for their constructive feedback during the review process.

Conflict of interest

The authors affirm that they have no known financial or interpersonal conflicts that would have appeared to have an impact on the research presented in this study.

Authors contribution

Aparnna Ravi: Method development, Coding, Data processing, Analysis, Visualization, Writing—original draft preparation, writing review and editing, Dhanyalekshmi Pillai: Conceptualization, Method development, writing review and editing, Christoph Gerbig: Data processing, Vishnu Thilakan: Analysis and Writing, Stephan Sitch: Model data providing, Sönke Zaehle: Writing & editing, Chandrashekhar Jha: EC flux tower data acquisition and processing and Writing & editing, Thara Anna: Writing & editing.

ORCID iDs

Aparnna Ravi https://orcid.org/0009-0007-2214-8427
Dhanyalekshmi Pillai https://orcid.org/0000-0002-8934-2140

References

Bazzi H *et al* 2024 Assimilating Sentinel-2 data in a modified vegetation photosynthesis and respiration model (VPRM) to improve the simulation of croplands CO_2 fluxes in Europe *Int. J. Appl. Earth Obs. Geoinf.* **127** 103666

Botía S *et al* 2022 The CO_2 record at the Amazon tall tower observatory: a new opportunity to study processes on seasonal and inter-annual scales *Glob. Change Biol.* **28** 588–611

Brown C F *et al* 2022 Dynamic world, near real-time global 10 m land use land cover mapping *Sci. Data* **9** 251

Burman P K D, Launiainen S, Mukherjee S, Chakraborty S, Gogoi N, Murkute C, Lohani P, Sarma D and Kumar K 2021 Ecosystem-atmosphere carbon and water exchanges of subtropical evergreen and deciduous forests in India *For. Ecol. Manage.* **495** 119371

Celesti M, van der Tol C, Cogliati S, Panigada C, Yang P, Pinto F, Rascher U, Miglietta F, Colombo R and Rossini M 2018 Exploring the physiological information of Sun-induced chlorophyll fluorescence through radiative transfer model inversion *Remote Sens. Environ.* **215** 97–108

Cervarich M *et al* 2016 The terrestrial carbon budget of South and Southeast Asia *Environ. Res. Lett.* **11** 105006

Chen A, Mao J, Ricciuto D, Lu D, Xiao J, Li X, Thornton P E and Knapp A K 2021 Seasonal changes in GPP/SIF ratios and their climatic determinants across the Northern Hemisphere *Glob. Change Biol.* **27** 5186–97

Chen F *et al* 2007 Description and evaluation of the characteristics of the NCAR high-resolution land data assimilation system *J. Appl. Meteorol. Climatol.* **46** 694–713

Dayalu A, Munger J W, Wofsy S C, Wang Y, Nehrkorn T, Zhao Y, McElroy M B, Nielsen C and Luus K 2017 VPRM-China: using the vegetation, photosynthesis and respiration model to partition contributions to CO_2 measurements in Northern China during the 2005–2009 growing seasons *Biogeosciences Discuss.* **2017** 1–29

De Bartolomeis P, Meterez A, Shu Z and Stocker B D 2023 An effective machine learning approach for predicting ecosystem CO_2 assimilation across space and time *EGUsphere* **2023** 1–31

Dou X, Yang Y and Luo J 2018 Estimating forest carbon fluxes using machine learning techniques based on eddy covariance measurements *Sustainability* **10** 203

Flexas J, Bota J, Galmés J, Medrano H and Ribas-Carbó M 2006 Keeping a positive carbon balance under adverse conditions: responses of photosynthesis and respiration to water stress *Physiol. Plant.* **127** 343–52

Frankenberg C *et al* 2011 New global observations of the terrestrial carbon cycle from GOSAT: patterns of plant fluorescence with gross primary productivity *Geophys. Res. Lett.* **38** L17706

Friedlingstein P *et al* 2022 Global carbon budget 2022 *Earth Syst. Sci. Data* **14** 4811–900

Friend A D *et al* 2007 FLUXNET and modelling the global carbon cycle *Glob. Change Biol.* **13** 610–33

Gerbig C and Koch F T 2021 Biosphere-atmosphere exchange fluxes for CO_2 from the vegetation photosynthesis and respiration model VPRM for 2006–2022 (https://doi.org/10.18160/VX78-HVA1)

Goroshi S, Singh R P, Pradhan R and Parihar J S 2014 Assessment of net primary productivity over India using Indian geostationary satellite (INSAT-3A) data *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **40** 561–8

Gourdji S M, Karion A, Lopez-Coto I, Ghosh S, Mueller K L, Zhou Y, Williams C A, Baker I T, Haynes K D and Whetstone J R 2022 A modified vegetation photosynthesis and respiration model (VPRM) for the Eastern USA and Canada, evaluated with comparison to atmospheric observations and other biospheric models *J. Geophys. Res. Biogeosci.* **127** e2021JG006290

Gray A N and Whittier T R 2014 Carbon stocks and changes on Pacific Northwest national forests and the role of disturbance, management and growth *For. Ecol. Manage.* **328** 167–78

Gu L, Han J, Wood J D, Chang C Y-Y and Sun Y 2019 Sun-induced Chl fluorescence and its importance for biophysical modeling of photosynthesis based on light reactions *New Phytol.* **223** 1179–91

Guanter L *et al* 2021 The TROPOSIF global Sun-induced fluorescence dataset from the Sentinel-5P TROPOMI mission *Earth Syst. Sci. Data* **13** 5423–40

Guanter L, Frankenberg C, Dudhia A, Lewis P E, Gómez-Dans J, Kuze A, Suto H and Grainger R G 2012 Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements *Remote Sens. Environ.* **121** 236–51

Guanter L *et al* 2014 Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence *Proc. Natl Acad. Sci.* **111** E1327–33

He B *et al* 2022 Worldwide impacts of atmospheric vapor pressure deficit on the interannual variability of terrestrial carbon sinks *Natl Sci. Rev.* **9** nwab150

Hersbach H *et al* 2020 The ERA5 global reanalysis *Q. J. R. Meteorol. Soc.* **146** 1999–2049

Ienco D, Interdonato R, Gaetano R and Minh D H T 2019 Combining Sentinel-1 and Sentinel-2 satellite image time series for land cover mapping via a multi-source deep learning architecture *ISPRS J. Photogramm. Remote Sens.* **158** 11–22

Jha C S, Thumaty K C, Rodda S R, Sonakia A and Dadhwal V K 2013 Analysis of carbon dioxide, water vapour and energy fluxes over an Indian teak mixed deciduous forest for winter and summer months using eddy covariance technique *J. Earth Syst. Sci.* **122** 1259–68

Jiao K, Liu Z, Wang W, Yu K, Mcgrath M J and Xu W 2024 Carbon cycle responses to climate change across China's terrestrial ecosystem: sensitivity and driving process *Sci. Total Environ.* **915** 170053

Jiye Z 2020 *A Data-Driven Upscale Product of Global Gross Primary Production, Net Ecosystem Exchange and Ecosystem Respiration* (National Institute for Environmental Studies)

Joiner J, Guanter L, Lindstrot R, Voigt M, Vasilkov A, Middleton E, Huemmrich K, Yoshida Y and Frankenberg C 2013 Global monitoring of terrestrial chlorophyll fluorescence from moderate spectral resolution near-infrared satellite measurements: methodology, simulations and application to GOME-2 *Atmos. Meas. Tech.* **6** 3883–930

Joiner J, Yoshida Y, Guanter L and Middleton E M 2016 New methods for the retrieval of chlorophyll red fluorescence from hyperspectral satellite instruments: simulations and application to GOME-2 and SCIAMACHY *Atmos. Meas. Tech.* **9** 3939–67

Joiner J, Yoshida Y, Zhang Y, Duveiller G, Jung M, Lyapustin A, Wang Y and Tucker C J 2018 Estimation of terrestrial global gross primary production (GPP) with satellite data-driven models and eddy covariance flux data *Remote Sens.* **10** 1346

Jung M, Henkel K, Herold M and Churkina G 2006 Exploiting synergies of global land cover products for carbon cycle modeling *Remote Sens. Environ.* **101** 534–53

Jung M *et al* 2020 Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach *Biogeosciences* **17** 1343–65

Kim J, Ryu Y, Dechant B, Lee H, Kim H S, Kornfeld A and Berry J A 2021 Solar-induced chlorophyll fluorescence is non-linearly related to canopy photosynthesis in a temperate evergreen needleleaf forest during the fall transition *Remote Sens. Environ.* **258** 112362

Köhler P, Frankenberg C, Magney T S, Guanter L, Joiner J and Landgraf J 2018 Global retrievals of solar-induced chlorophyll fluorescence with TROPOMI: first results and intersensor comparison to OCO-2 *Geophys. Res. Lett.* **45** 10456–63

Köhler P, Guanter L and Joiner J 2015 A linear method for the retrieval of Sun-induced chlorophyll fluorescence from GOME-2 and SCIAMACHY data *Atmos. Meas. Tech.* **8** 2589–608

Kountouris P, Gerbig C, Rödenbeck C, Karstens U, Koch T F and Heimann M 2018 Atmospheric CO_2 inversions on the mesoscale using data-driven prior uncertainties: quantification of the European terrestrial CO_2 fluxes *Atmos. Chem. Phys.* **18** 3047–64

Kunik L, Bowling D R, Raczka B, Frankenberg C, Köhler P, Cheng R, Smith K R, Goulden M, Jung M and Lin J C 2023 Satellite-based solar-induced fluorescence tracks seasonal and elevational patterns of photosynthesis in California's Sierra Nevada mountains *Environ. Res. Lett.* **19** 014008

Li X, Wang Y-P, Lu X and Yan J 2021 Diagnosing the impacts of climate extremes on the interannual variations of carbon fluxes of a subtropical evergreen mixed forest *Agric. For. Meteorol.* **307** 108507

Li X and Xiao J 2019a A global, 0.05-degree product of solar-induced chlorophyll fluorescence derived from OCO-2, MODIS and reanalysis data *Remote Sens.* **11** 517

Li X and Xiao J 2019b Mapping photosynthesis solely from solar-induced chlorophyll fluorescence: a global, fine-resolution dataset of gross primary production derived from OCO-2 *Remote Sens.* **11** 2563

Li X *et al* 2018 Solar-induced chlorophyll fluorescence is strongly correlated with terrestrial photosynthesis for a wide variety of biomes: first global analysis based on OCO-2 and flux tower observations *Glob. Change Biol.* **24** 3990–4008

Liu H, Liu J, Yin Y, Walther S, Ma X, Zhang Z and Chen Y 2023 Improved vegetation photosynthetic phenology monitoring in the northern ecosystems using total canopy solar-induced chlorophyll fluorescence derived from TROPOMI *J. Geophys. Res. Biogeosci.* **128** e2022JG007369

Liu X, Liu L, Hu J, Guo J and Du S 2020 Improving the potential of red SIF for estimating GPP by downscaling from the canopy level to the photosystem level *Agric. For. Meteorol.* **281** 107846

Liu Z, Shao Q and Liu J 2014 The performances of MODIS-GPP and-ET products in China and their sensitivity to input data (FPAR/LAI) *Remote Sens.* **7** 135–52

Lu H *et al* 2021 Comparing machine learning-derived global estimates of soil respiration and its components with those from terrestrial ecosystem models *Environ. Res. Lett.* **16** 054048

Luus K A and Lin J C 2015 The polar vegetation photosynthesis and respiration model (PolarVPRM): a parsimonious, satellite data-driven model of high-latitude CO_2 exchange *Geosci. Model Dev. Discuss.* **8** 979–1027

Magney T S *et al* 2019 Disentangling changes in the spectral shape of chlorophyll fluorescence: implications for remote sensing of photosynthesis *J. Geophys. Res. Biogeosci.* **124** 1491–507

Mahadevan P, Wofsy S C, Matross D M, Xiao X, Dunn A L, Lin J C, Gerbig C, Munger J W, Chow V Y and Gottlieb E W 2008 A satellite-based biosphere parameterization for net ecosystem CO_2 exchange: vegetation photosynthesis and respiration model (VPRM) *Glob. Biogeochem. Cycles* **22** GB2005

Martens B, Miralles D G, Lievens H, van der Schalie R, de Jeu R A, Fernández-Prieto D, Beck H E, Dorigo W A and Verhoest N E C 2017 GLEAM v3: satellite-based land evaporation and root-zone soil moisture *Geosci. Model Dev.* **10** 1903–25

McKinley D C *et al* 2011 A synthesis of current knowledge on forests and carbon storage in the United States *Ecol. Appl.* **21** 1902–24

Meir P, Metcalfe D B, Costa A and Fisher R A 2008 The fate of assimilated carbon during drought: impacts on respiration in Amazon rainforests *Phil. Trans. R. Soc.* B **363** 1849–55

Melaas E K, Richardson A D, Friedl M A, Dragoni D, Gough C M, Herbst M, Montagnani L and Moors E 2013 Using FLUXNET data to improve models of springtime vegetation activity onset in forest ecosystems *Agric. For. Meteorol.* **171** 46–56

Molchanov A 2009 Effect of moisture availability on photosynthetic productivity and autotrophic respiration of an oak stand *Russ. J. Plant Physiol.* **56** 769–79

Nayak R, Patel N and Dadhwal V 2013 Inter-annual variability and climate control of terrestrial net primary productivity over India *Int. J. Climatol.* **33** 132–42

Nayak R, Patel N and Dadhwal V 2015 Spatio-temporal variability of net ecosystem productivity over India and its relationship to climatic variables *Environ. Earth Sci.* **74** 1743–53

Nayak R K, Patel N and Dadhwal V 2010 Estimation and analysis of terrestrial net primary productivity over India by remote-sensing-driven terrestrial biosphere model *Environ. Monit. Assess.* **170** 195–213

Parazoo N C *et al* 2018 Spring photosynthetic onset and net CO_2 uptake in Alaska triggered by landscape thawing *Glob. Change Biol.* **24** 3416–35

Parazoo N C, Frankenberg C, Köhler P, Joiner J, Yoshida Y, Magney T, Sun Y and Yadav V 2019 Towards a harmonized long-term spaceborne record of far-red solar-induced fluorescence *J. Geophys. Res. Biogeosci.* **124** 2518–39

Patra P *et al* 2013 The carbon budget of South Asia *Biogeosciences* **10** 513–27

Patra P, Niwa Y, Schuck T, Brenninkmeijer C, Machida T, Matsueda H and Sawa Y 2011 Carbon balance of South Asia constrained by passenger aircraft CO_2 measurements *Atmos. Chem. Phys.* **11** 4163–75

Paul-Limoges E, Damm A, Hueni A, Liebisch F, Eugster W, Schaepman M E and Buchmann N 2018 Effect of environmental conditions on Sun-induced fluorescence in a mixed forest and a cropland *Remote Sens. Environ.* **219** 310–23

Peters W *et al* 2007 An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker *Proc. Natl Acad. Sci.* **104** 18925–30

Peylin P *et al* 2013 Global atmospheric carbon budget: results from an ensemble of atmospheric CO_2 inversions *Biogeosciences* **10** 6699–720

Potter C S, Randerson J T, Field C B, Matson P A, Vitousek P M, Mooney H A and Klooster S A 1993 Terrestrial ecosystem production: a process model based on global satellite and surface data *Glob. Biogeochem. Cycles* **7** 811–41

Qiu R, Han G, Ma X, Xu H, Shi T and Zhang M 2020 A comparison of OCO-2 SIF, MODIS GPP and GOSIF data from gross primary production (GPP) estimation and seasonal cycles in North America *Remote Sens.* **12** 258

Rao A S, Bala G, Ravindranath N and Nemani R 2019 Multi-model assessment of trends, variability and drivers of terrestrial carbon uptake in India *J. Earth Syst. Sci.* **128** 1–19

Ravi A, Pillai D, Thilakan V and Mathew T A 2024 Methodological advancement in deriving primary productivity and ecosystem respiration fluxes across different biomes *MethodsX* **12** 102773

Rodda S R, Thumaty K C, Jha C S and Dadhwal V K 2016 Seasonal variations of carbon dioxide, water vapor and energy fluxes in tropical Indian mangroves *Forests* **7** 35

Rodda S R, Thumaty K C, Praveen M, Jha C S and Dadhwal V K 2021 Multi-year eddy covariance measurements of net ecosystem exchange in tropical dry deciduous forest of India *Agric. For. Meteorol.* **301** 108351

Rodgers C D 2000 *Inverse Methods for Atmospheric Sounding: Theory and Practice* vol 2 (World Scientific)

Sarma D, Burman P K D, Chakraborty S, Gogoi N, Bora A, Metya A, Datye A, Murkute C and Karipot A 2022 Quantifying the net ecosystem exchange at a semi-deciduous forest in Northeast India from intra-seasonal to the seasonal time scale *Agric. For. Meteorol.* **314** 108786

Sitch S *et al* 2015 Recent trends and drivers of regional sources and sinks of carbon dioxide *Biogeosciences* **12** 653–79

Sitch S et al 2008 Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five dynamic global vegetation models (DGVMs) *Glob. Change Biol.* **14** 2015–39
Smith W et al 2018 Chlorophyll fluorescence better captures seasonal and interannual gross primary productivity dynamics across dryland ecosystems of southwestern North America *Geophys. Res. Lett.* **45** 748–57
Sun Y, Frankenberg C, Jung M, Joiner J, Guanter L, Köhler P and Magney T 2018 Overview of solar-induced chlorophyll fluorescence (SIF) from the Orbiting Carbon Observatory-2: retrieval, cross-mission comparison and global monitoring for GPP *Remote Sens. Environ.* **209** 808–23
Sun Y et al 2017 OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence *Science* **358** eaam5747
Taylor T E et al 2020 OCO-3 early mission operations and initial (vEarly) XCO_2 and SIF retrievals *Remote Sens. Environ.* **251** 112032
Thilakan V, Pillai D, Gerbig C, Galkowski M, Ravi A and Anna Mathew T 2022 Towards monitoring the CO_2 source–sink distribution over India via inverse modelling: quantifying the fine-scale spatiotemporal variability in the atmospheric CO_2 mole fraction *Atmos. Chem. Phys.* **22** 15287–312
Thompson R L et al 2016 Top–down assessment of the Asian carbon budget since the mid 1990s *Nat. Commun.* **7** 10724
Turner D P et al 2006 Evaluation of MODIS NPP and GPP products across multiple biomes *Remote Sens. Environ.* **102** 282–92
Upton S, Reichstein M, Gans F, Peters W, Kraft B and Bastos A 2024 Constraining biospheric carbon dioxide fluxes by combined top-down and bottom-up approaches *Atmos. Chem. Phys.* **24** 2555–82
van Dijke A J H, Mallick K, Schlerf M, Machwitz M, Herold M and Teuling A J 2020 Examining the link between vegetation leaf area and land–atmosphere exchange of water, energy and carbon fluxes using FLUXNET data *Biogeosciences* **17** 4443–57
van der Meer P, Jorritsma I and Kramer K 2002 Assessing climate change effects on long-term forest development: adjusting growth, phenology and seed production in a gap model *For. Ecol. Manage.* **162** 39–52
Valsala V, Tiwari Y K, Pillai P, Roxy M, Maksyutov S and Murtugudde R 2013 Intraseasonal variability of terrestrial biospheric CO_2 fluxes over India during summer monsoons *J. Geophys. Res. Biogeosci.* **118** 752–69
Varghese R and Behera M 2019 Annual and seasonal variations in gross primary productivity across the agro-climatic regions in India *Environ. Monit. Assess.* **191** 631
Virkkala A M et al 2021 Statistical upscaling of ecosystem CO_2 fluxes across the terrestrial tundra and boreal domain: regional patterns and uncertainties *Glob. Change Biol.* **27** 4040–59
Wang X, Chen J M and Ju W 2020 Photochemical reflectance index (PRI) can be used to improve the relationship between gross primary productivity (GPP) and Sun-induced chlorophyll fluorescence (SIF) *Remote Sens. Environ.* **246** 111888
Wood J D, Griffis T J, Baker J M, Frankenberg C, Verma M and Yuen K 2017 Multiscale analyses of solar-induced florescence and gross primary production *Geophys. Res. Lett.* **44** 533–41
Yu G-R et al 2013 Spatial patterns and climate drivers of carbon fluxes in terrestrial ecosystems of China *Glob. Change Biol.* **19** 798–810
Yu L, Wen J, Chang C, Frankenberg C and Sun Y 2019 High-resolution global contiguous SIF of OCO-2 *Geophys. Res. Lett.* **46** 1449–58
Zhang F, Chen J M, Chen J, Gough C M, Martin T A and Dragoni D 2012 Evaluating spatial and temporal patterns of MODIS GPP over the conterminous US against flux measurements and a process model *Remote Sens. Environ.* **124** 717–29
Zhang Y, Guanter L, Berry J A, van der Tol C, Yang X, Tang J and Zhang F 2016 Model-based analysis of the relationship between Sun-induced chlorophyll fluorescence and gross primary production for remote sensing applications *Remote Sens. Environ.* **187** 145–55
Zhang Y, Yu Q, Jiang J and Tang Y 2008 Calibration of Terra/MODIS gross primary production over an irrigated cropland on the North China Plain and an alpine meadow on the Tibetan Plateau *Glob. Change Biol.* **14** 757–67
Zhang Z, Chen J M, Guanter L, He L and Zhang Y 2019 From canopy-leaving to total canopy far-red fluorescence emission for remote sensing of photosynthesis: first results from TROPOMI *Geophys. Res. Lett.* **46** 12030–40
Zhang Z et al 2020 Reduction of structural impacts and distinction of photosynthetic pathways in a global estimation of GPP from space-borne solar-induced chlorophyll fluorescence *Remote Sens. Environ.* **240** 111722
Zhao F, Ma W, Zhao J, Guo Y, Tariq M and Li J 2024 Global retrieval of the spectrum of terrestrial chlorophyll fluorescence: first results with TROPOMI *Remote Sens. Environ.* **300** 113903
Zheng B, Chevallier F, Ciais P, Broquet G, Wang Y, Lian J and Zhao Y 2020 Observing carbon dioxide emissions over China's cities and industrial areas with the orbiting carbon Observatory-2 *Atmos. Chem. Phys.* **20** 8501–10
Zheng H, Lin H, Zhou W, Bao H, Zhu X, Jin Z, Song Y, Wang Y, Liu W and Tang Y 2019 Revegetation has increased ecosystem water-use efficiency during 2000–2014 in the Chinese Loess Plateau: evidence from satellite data *Ecol. Indic.* **102** 507–18