



Global GPP from microwave remote sensing–driven dynamics: feasibility, strengths, and limitations

Ruxandra-Maria Zotta¹, Moritz Clemens Müller¹, Raul Lezameta², Sophia Walther³, Matthias Forkel⁴, and Wouter Dorigo¹

¹Department of Geodesy and Geoinformation, TU Wien, Wiedner Hauptstraße 8–10, 1040 Vienna, Austria

²Environmental Earth Observation Information Technology GmbH, Fürstenweg 176, Innsbruck 6020, Austria

³Max Planck Institute for Biogeochemistry, Beutenberg Campus, Hans-Knöll-Straße 10, 07745 Jena, Germany

⁴Technische Universität Dresden, Faculty of Environmental Sciences, Helmholtzstr. 10, 01069 Dresden Germany

Correspondence: ruxandra-maria.zotta@tuwien.ac.at

Abstract.

Long-term monitoring of gross primary production (GPP) is essential for quantifying terrestrial carbon uptake and ecosystem responses to climate variability and extremes. VODCA2GPP is a data-driven upscaling framework that links eddy-covariance (EC) GPP observations to passive-microwave vegetation optical depth (VOD), together with reanalysis air temperature, providing a multi-decadal microwave-based complement to optical GPP products. However, its reliance on reanalysis temperature reduces the methodological independence of the microwave-based estimate, and its sensitivity to water stress is limited in some ecosystems. Here, we develop and evaluate an updated microwave-driven GPP framework in which the time-varying environmental drivers are derived from microwave observations. At the same time, static land-cover information provides biome context. The framework produces 8-day GPP at 0.25° spatial resolution over 2002–2020. Using a sequence of Random Forest (RF) models trained to upscale EC GPP, we test the incremental contribution of different predictor groups. We replace ERA5-Land 2-m air temperature (T2M) with land-surface temperature (LST) derived from harmonised Ka-band brightness temperatures from AMSR-E, AMSR2, and SSMIS F17, and then evaluate land-cover (LC) fractions as biome context, a spatially varying multi-year mean of L-band VOD as persistent canopy-structure information, and surface soil moisture (SSM) and root-zone soil moisture (RZSM) as shallow and deeper water-availability constraints. The models are compared with external GPP products, sun-induced chlorophyll fluorescence (SIF), and drought-related datasets.

Predictive skill is assessed using site-grouped 10-fold cross-validation, keeping all observations from a site together in either the training or validation set within each fold. Replacing reanalysis T2M with microwave-derived LST yielded broadly comparable site-level predictive performance. Adding LC fractions produces the largest individual improvement in predictive skill (median r : 0.690 to 0.762; RMSE: 2.646 to 2.345 g C m⁻² d⁻¹), while the spatially varying multi-year L-band VOD mean provides a smaller additional improvement by adding persistent canopy-structure information. Structural and hydrological constraints improve agreement with external GPP products, SIF, and drought indicators, with the clearest gains in short-vegetation and water-limited regions; among the hydrological predictors, RZSM provides the strongest improvement.



25 The best-performing configuration achieves a median site-level correlation of $r = 0.798$ and an RMSE of $2.294 \text{ g C m}^{-2} \text{ d}^{-1}$. Despite these improvements, the GPP magnitude of our models remains systematically higher than in the comparison products.

Overall, the results show that the updated framework can reproduce long-term GPP dynamics without reanalysis-based temperature forcing, while static LC fractions provide complementary biome context. Among the added predictors, LC fractions provide the largest overall gain in predictive skill and help distinguish productivity among vegetation types, the multi-year L-
 30 band VOD mean adds persistent canopy-structure information, and RZSM improves the representation of water-limited GPP variability, especially at short-vegetation sites. Remaining limitations include the mismatch between EC footprints and the much coarser satellite grid, uneven environmental and geographic coverage of the tower network, and environmental controls on GPP that are not explicitly represented by the current predictor set.

1 Introduction

35 Gross primary production (GPP) is the total uptake of atmospheric CO_2 by terrestrial vegetation through photosynthesis and, at the ecosystem scale, is commonly expressed as the sum of net primary production and autotrophic (plant) respiration (Roxburgh et al., 2005; Collalti and Prentice, 2019). As the largest gross CO_2 flux on land, GPP underpins ecosystem functioning and the strength of the terrestrial carbon sink to partially offset anthropogenic emissions (Beer et al., 2010; Friedlingstein et al., 2024). Because photosynthesis and plant respiration are constrained by water and energy, GPP also mediates carbon–water
 40 coupling; therefore, sustained monitoring of GPP is essential to diagnose how heat, and soil-moisture anomalies modulate photosynthetic activity and the strength of the terrestrial carbon sink, enabling early detection of stress and a more robust attribution of land–atmosphere feedbacks to hydro-meteorological drivers (Beer et al., 2010; Jung et al., 2017; Humphrey et al., 2021). Eddy-covariance (EC) observations provide the primary in situ constraint on ecosystem GPP and underpin both process-model evaluation and data-driven upscaling (Baldocchi, 2014; Pastorello et al., 2020). However, the spatially sparse
 45 and uneven distribution of flux towers precludes direct global monitoring from EC measurements alone.

At the global scale, optical Earth observation data, combined with (semi-)empirical or machine-learning models, have been the primary means of estimating GPP (O’Sullivan et al. (2020); Jung et al. (2020)). The most widely used are light-use-efficiency (LUE) frameworks (e.g., Running et al. (2004)), which use satellite canopy-state products, such as fraction of absorbed photosynthetically active radiation (fAPAR) and leaf area index (LAI), sometimes complemented by vegetation indices,
 50 with biome-specific parameters and meteorological scalars (Zhu et al., 2024). In these models, absorbed light is typically estimated from incoming PAR and a satellite-derived absorption term (fAPAR), while biome-dependent lookup parameters, typically derived from literature and calibrated against ecosystem observations, define the maximum LUE (Running et al., 2004). Meteorological scalars then down-regulate this potential efficiency under suboptimal conditions, most commonly using minimum temperature and atmospheric dryness (vapour pressure deficit; VPD), sometimes complemented by water-stress proxies,
 55 to better capture day-to-day limitations on photosynthesis (Running and Zhao, 2019). Increasingly, sun-induced fluorescence (SIF) is also used as an additional constraint, as it tracks changes in photosynthetic activity more directly than canopy greenness alone (Schickling et al., 2016). Machine-learning upscaling of tower fluxes using optical remote sensing and meteorological



predictors (FLUXCOM; Tramontana et al. (2016); Jung et al. (2020); Nelson & Walther et al. (2024)) has delivered global GPP estimates widely used for model benchmarking and diagnosing variability (Jung et al. (2020)). Nonetheless, optical methods can suffer from cloud and aerosol contamination, illumination effects, and saturation in dense canopies, which degrade coverage and sensitivity precisely in high-productivity, persistently cloudy regions (Turner et al. (2006)). These limitations motivate complementary observing systems that provide all-weather capability and long-term continuity while remaining anchored to EC constraints.

Compared to observations from the optical domain, microwave remote sensing is much less affected by weather conditions and does not depend on illumination. Microwave vegetation optical depth (VOD) retrievals are nevertheless not limitation-free: radio-frequency interference can contaminate individual bands, while frozen or snow-covered surface conditions require screening because they can produce unreliable retrievals (Zotta et al., 2024a). VOD quantifies the attenuation of microwave radiation emitted or reflected by the Earth's surface, due to the vegetation layer, and provides a combined proxy for canopy water content and above-ground biomass that complements optical data (Frappart et al. (2020); Schmidt et al. (2022)). Two complementary physical pathways motivate VOD as a predictor of GPP. First, VOD relates to canopy structural properties, including LAI and foliage biomass (Momen et al. (2017); Yang et al. (2025)), which determine the amount of photosynthetically active radiation a canopy can absorb and therefore constrain its photosynthetic capacity, analogous to the role of LAI/fAPAR in optical LUE models. Second, VOD is sensitive to vegetation water content and can therefore capture changes in canopy hydration associated with plant water stress. Because declining plant water status induces stomatal regulation and limits CO₂ uptake, this provides an indirect pathway linking VOD variability to photosynthetic activity, particularly under water-limited conditions (Konings and Gentine, 2017; Martens et al., 2017; Zhang et al., 2019). This second pathway is, however, more difficult to isolate because the water-status signal is superimposed on VOD's strong sensitivity to vegetation biomass and canopy structure (Momen et al., 2017; Konings et al., 2021).

VOD observations at higher microwave frequencies (e.g., C-, X-, and Ku-band) have been shown to strongly covary with ecosystem productivity at sub-monthly scales, across biomes, although this relationship varies with land cover (LC), sensor and signal characteristics, and can exhibit a seasonal hysteresis (Teubner et al. (2018)), consistent with VOD's dual sensitivity to structure and water status. Building on this insight, Teubner et al. (2019) introduced a carbon-sink-driven (i.e., accumulation of biomass between two consecutive time steps) framework for estimating GPP from VOD by upscaling FLUXNET EC data, positioning the approach as a complement to "source-driven" optical LUE models. Teubner et al. (2020) extended the model to include an explicit air temperature term from ERA5 reanalysis to model the temperature dependence of autotrophic respiration. These methodological developments, together with the launch of the long-term VOD Climate Archive (VODCA; Moesinger et al. (2020); Zotta et al. (2024b)), led to the creation of VODCA2GPP (Wild et al. (2022)), the first long-term (1988-2020), microwave-derived GPP product. VODCA2GPP was compared with GPP products from optical remote sensing and dynamic global vegetation models, and showed good correspondence in both temporal and spatial domains, though agreement varied across biomes, with tropical and arid regions showing the poorest agreement (Wild et al., 2022).

The limitations of VODCA2GPP point to three opportunities to improve a microwave-driven GPP framework. First, Ka-band passive microwave observations can provide a land-surface temperature proxy that is closely coupled to surface thermal



conditions and may therefore replace reanalysis 2-m air temperature (T2M) as the time-varying temperature constraint (Holmes et al., 2009; Good et al., 2017). Second, VOD reflects vegetation water content, biomass, and canopy structure, but drought-induced reductions in GPP can occur via stomatal regulation without proportional changes in canopy water content. Because this coupling varies with plant hydraulic strategy along the iso- to anisohydric continuum, we complement VOD with plant-accessible soil water (Konings and Gentile, 2017; Konings et al., 2019; Zhang et al., 2019). Root-zone soil moisture (RZSM) is especially relevant because it integrates water availability over longer timescales and depths than surface soil moisture (SSM). Third, the relationship between dynamic microwave signals and GPP depends on ecosystem structure and vegetation type. While higher-frequency VOD already contains information on vegetation biomass and water content, L-band VOD provides a complementary structural constraint because its longer wavelength is more sensitive to woody vegetation and is less prone to saturation in dense canopies. Together with LC information, it can therefore provide additional context on persistent vegetation structure and biome differences within a globally trained model (Rodríguez-Fernández et al., 2018; Mialon et al., 2020).

This study assesses the feasibility, strengths, and limitations of a global GPP framework whose time-varying environmental drivers are derived from microwave observation. The framework produces 8-daily GPP at 0.25° spatial resolution over 2002–2020. We first test whether Ka-band-derived daytime and nighttime LST can replace the ERA5-Land T2M thermal constraint while retaining comparable predictive performance. Because several existing GPP products also ingest ERA5(-Land), this replacement additionally reduces one source of shared reanalysis forcing across products (e.g., Nelson & Walther et al. (2024); Yuan et al. (2025)). We then quantify the added value of structural and biome information (L-VOD and LC) and hydrological constraints (SSM and RZSM), with particular emphasis on water-limited vegetation, where GPP responses to soil-water availability remain difficult to constrain consistently across products (Green et al., 2019) and constitutes a weakness of the VODCA2GPP implementation (Wild et al., 2022). Finally, we evaluate GPP magnitude and temporal agreement with external products, water-related variability, and sensitivity to the composition of the flux-tower training network. Overall, we assess the potential of a microwave-driven framework to reproduce GPP dynamics across ecosystems, while identifying the remaining limitations in spatial extrapolation, environmental constraints, and absolute GPP magnitude.

2 Data

2.1 Model input data

The GPP framework is developed for 2002–2020 at an 8-day temporal resolution on a common 0.25° spatial grid. The predictor set combines information on vegetation state and structure, surface thermal conditions, water availability, and vegetation composition. Predictors are harmonised to the common 8-day temporal resolution and 0.25° grid. Table 1 summarises the predictors used in the GPP models.



Table 1. Predictors used in the GPP models, grouped by functional role. All predictors were aggregated to the common 0.25° grid and 8-daily temporal resolution (if dynamic).

Predictor [unit]	Source	Band / frequency	Representation in model
<i>VOD-derived vegetation predictors</i>			
VOD [-]	VODCA v2, Ku band	Ku (~19 GHz)	Dynamic canopy water and vegetation-state signal; daily observations aggregated to 8-day means and smoothed with an 11-step Savitzky–Golay filter.
dVOD [-]	Derived from VOD	Ku (~19 GHz)	Change in vegetation state; finite difference between consecutive 8-day VOD composites, followed by the same smoothing as VOD.
medVOD [-]	Derived from VOD	Ku (~19 GHz)	Static background vegetation state; temporal median VOD over the model period.
<i>Thermal predictors</i>			
LST_{day} [K]	AMSR-E, AMSR2, SSMIS F17	Ka (~37 GHz)	Daytime thermal constraint on photosynthesis, capturing temperature effects on plant metabolism and heat stress together with coupling to stomatal regulation, transpiration, and evaporative cooling.
LST_{night} [K]	AMSR-E, AMSR2, SSMIS F17	Ka (~37 GHz)	Complementary nighttime thermal constraint; although photosynthesis is absent, nighttime temperature affects ecosystem respiration and plant metabolic and developmental processes that influence subsequent carbon uptake.
T2M [K]	ERA5-Land	not microwave-derived	2-m air temperature used as the reanalysis-based benchmark thermal predictor.
<i>Structural and compositional predictors</i>			
L-VOD [-]	VODCA L; SMOS and SMAP	L (~1.4 GHz)	Persistent vegetation structure and woody biomass; multi-year 2010–2020 mean used as a static predictor.
LC fractions [-] (0, 1)	ESA CCI Land Cover	not microwave-derived	Static vegetation composition; 300-m classes converted to 11 PFT fractions, aggregated to 0.25°, and averaged over 2002–2020.
<i>Hydrological predictors</i>			
SSM [$\text{m}^3 \text{m}^{-3}$]	ESA CCI Soil Moisture GAPFILLED	Multi-sensor passive (L/C/X/Ku) and active (C) ~ 1.4–19 GHz	Rapidly varying near-surface (~0–5 cm) water-state constraint, capturing short-term wetting and drying that can limit stomatal conductance and photosynthesis, particularly in shallow-rooted vegetation.
RZSM [mm]	G3P / C3S-based RZSM	Multi-sensor (passive L/C/X/Ku) microwave SSM input (~ 1.4–19 GHz)	Deeper (~ 10 - 100 cm plant-accessible water-storage constraint with longer hydrological memory, representing sustained water availability that regulates stomatal conductance and photosynthesis during soil drying and drought.

2.1.1 VOD and VOD-based metrics

We use VOD from the VODCA v2 dataset (Zotta et al., 2024b, a), which provides long-term, homogenised records of passive microwave-derived VOD retrieved with the Land Parameter Retrieval Model (LPRM; Owe et al. (2002); Meesters et al. (2005); van Der Schalie et al. (2017)). In this study, we use the Ku-band VOD product (18.7 GHz), rather than VODCA CXKu (fusing C-, X- and Ku-band) used by Wild et al. (2022), because Ku is less affected by radio-frequency interference than C- and X-band, especially in Europe and the USA (de Nijs et al., 2015), where most of the EC towers are located. While C-, X-, and Ku-band VOD observations are highly correlated (Zotta et al., 2024b), the shorter wavelength of Ku-band generally reduces canopy penetration and increases its relative sensitivity to upper-canopy vegetation compared with lower-frequency VOD (Schmidt et al., 2022). For the present analysis, the Ku-band predictor is represented on the common 0.25° grid and used over the 2002–2020 study period.



Similar to Wild et al. (2022) and Teubner et al. (2019), we use three complementary Ku-band VOD-based predictors for each grid cell at an 8-day cadence. First, daily observations are grouped onto the model's 8-day time step to reduce day-to-day retrieval noise; each 8-day composite is computed using the mean of the valid observations within that interval. The bulk VOD composite, therefore, represents canopy state over an 8-day interval rather than an instantaneous observation. Second, the short-term change $dVOD$, computed as the finite difference between consecutive 8-day composites, summarises rapid increases or decreases in canopy water content associated with leaf growth or senescence, disturbance, or acute water stress. Third, a time-invariant $medVOD$ predictor is calculated for each grid cell using the same common model-training period, rather than a tower-specific period, and is defined as the temporal median of Ku-band VOD. Because $medVOD$ is time-invariant, it cannot explicitly follow persistent land-cover change such as deforestation; in rapidly changing pixels, it may therefore retain structural information from earlier conditions. Dynamic VOD and $dVOD$ can respond to such changes, but this limitation of $medVOD$ should be acknowledged. We use a Savitzky-Golay filter (11 x 8-day steps, approximately 88 days) to reduce high-frequency observational noise. Gap filling is not performed.

To add a biomass-sensitive structural descriptor, we compute for each 0.25° pixel the multi-year mean of VODCA L (Zotta et al., 2024b). The resulting predictor is time-invariant within a pixel but varies spatially among pixels and leverages L-band's established sensitivity to the woody vegetation layer and its reduced saturation in dense canopies relative to higher microwave frequencies (Rodríguez-Fernández et al. (2018)). We use a multi-year mean rather than time-varying observations, because L-band VOD records begin only in 2010, whereas the GPP framework proposed here starts in 2002. The Ku-based $medVOD$ retains the median statistic used in the original framework to mitigate residual outliers, whereas L-VOD is used solely as a climatological structural descriptor over its shorter record and is therefore summarised by its temporal mean. The joint use of L-band VOD with dynamic Ku-band VOD is intended to separate persistent canopy structure from faster canopy-water and phenological variability. We henceforth refer to VODCA Ku VOD simply as VOD and to VODCA L as L-VOD.

2.1.2 LST derived from Ka-band brightness temperatures

We derive separate daytime and nighttime thermal predictors, LST_{day} and LST_{night} , from vertically polarised Ka-band brightness temperatures. Retaining both observations provides complementary information on the diurnal surface thermal regime, which is related to vegetation energy balance and carbon exchange (Still et al., 2021; Li et al., 2021b; Liu et al., 2024).

To construct a continuous 2002–2020 record, we combine observations from AMSR-E on Aqua (36.5 GHz; $\approx 14 \times 8$ km; 2002–2011; Kawanishi et al., 2003; Knowles et al., 2006), AMSR2 on GCOM-W1 (36.5 GHz; $\approx 12 \times 7$ km; since 2012; Meier et al., 2018), and SSMIS F17 (37.0 GHz; $\approx 37 \times 28$ – 38×30 km; Berg et al., 2018). AMSR-E and AMSR2 provide the main temporal record, while SSMIS F17 is used over 2008–2015 as a common reference spanning the transition between the two AMSR sensors. All the brightness temperature observations were processed within the ESA CCI Soil Moisture project.

The 36.5/37-GHz V-pol channels are used because brightness temperature at these frequencies is highly sensitive to LST, and we convert the harmonised observations to LST using the empirical relationship of Holmes et al. (2009). Retrieval performance is strongest over vegetated surfaces and becomes less robust over sparse or bare surfaces, where surface-emissivity variability has a larger influence.



The sensors differ in local observation time. AMSR-E and AMSR2 cross the Equator at approximately 13:30/01:30 local time for the ascending/descending overpasses, compared with approximately 18:20/06:00 for SSMIS F17 (NSIDC, 2019; Remote Sensing Systems, 2025; Boston College Institute for Scientific Research, 2025). Ascending and descending observations are used to construct LST_{day} and LST_{night} , respectively. Because F17 samples different phases of the diurnal temperature cycle and DMSP platforms are affected by orbital drift (Yang et al., 2011), differences in observation time remain a source of uncertainty that cannot be fully removed by inter-sensor harmonisation.

AMSR-E and AMSR2 brightness temperatures are harmonised to SSMIS F17 using per-pixel cumulative distribution function (CDF) matching, separately for daytime and nighttime observations (Liu et al., 2011; Moesinger et al., 2020). Transfer functions are derived from same-day observations during sensor-overlap periods, with at least 20 observation pairs required per grid cell, and subsequently applied to the full AMSR records. Multiple harmonised observations on the same day are averaged.

Finally, the harmonised brightness temperatures are converted to LST using the empirical relationships of Holmes et al. (2009):

$$LST_{\text{day}} = 0.898 T_{\text{B,day},36.5\text{V}} + 44.2, \quad (1)$$

$$LST_{\text{night}} = 0.893 T_{\text{B,night},36.5\text{V}} + 44.8. \quad (2)$$

2.1.3 Surface and root zone soil moisture

We use two complementary microwave-based soil moisture constraints representing different depths and hydrological timescales. Surface soil moisture (SSM; $\text{m}^3 \text{m}^{-3}$) is taken from the ESA CCI Soil Moisture GAFILLED climate data record (Preimesberger et al., 2025, 2024), which combines active and passive microwave observations into a daily global record on a 0.25° grid and represents approximately the upper 0–5 cm of the soil (Dorigo et al., 2017). We use the GAFILLED product because the standard ESA CCI Soil Moisture record contains persistent data gaps in densely vegetated regions, particularly in tropical forests, where microwave attenuation limits the retrieval of SM (Preimesberger et al., 2025). Gapfilling is done using DCT-PLS interpolation of the available SSM observations, with additional C- and L-band retrievals used as anchors in densely vegetated regions where microwave attenuation causes persistent data gaps (Preimesberger et al., 2025). Retrieval uncertainty nevertheless remains higher under dense vegetation. For non-frozen conditions, the gap filling is based on SM observations rather than ancillary meteorological or vegetation predictors; frozen periods are treated by temporal interpolation.

To represent deeper soil-water availability, we use root-zone soil moisture (RZSM) from the G3P product (Güntner et al., 2026). G3P derives RZSM from the C3S Soil Moisture COMBINED record (Dorigo et al., 2018) using the exponential-filter framework of Pasik et al. (2023). SSM and RZSM therefore share a common satellite SM heritage but differ in product version and subsequent processing. The exponential filter produces progressively smoother and delayed estimates with increasing depth and was calibrated against International Soil Moisture Network observations for the 0–10, 10–40, 0–100, and 100–200 cm layers (Pasik et al., 2023).



For the GPP models, we use RZSM integrated over 10–100 cm, excluding the 0–10 cm layer to reduce redundancy with the separate near-surface SSM predictor. Water storage is calculated from the 10–40 and 40–100 cm layers as

$$RZSM_{10-100} = 300\theta_{10-40} + 600\theta_{40-100}, \quad (3)$$

200 where 300 and 600 mm are the respective layer thicknesses and θ denotes volumetric SM. The resulting $RZSM_{10-100}$ is expressed in millimetres of water. Compared with SSM, this metric suppresses short-lived surface fluctuations and retains the longer hydrological memory of the rooting zone (Fan et al., 2017; Li et al., 2021b).

Both daily SSM and RZSM are aggregated to 8-day means. For each variable, we additionally derive an anomaly by removing the climatological seasonal cycle.

205 2.1.4 Land cover

Land-cover information is derived from the ESA CCI Land Cover time series (v2.0.7, 1992–2015; v2.1.1, 2016–2022), available annually at 300 m resolution (Defourny and ESA Land Cover CCI Project Team, 2017). For model input, the native classes are converted to fractional coverages of 11 plant functional types (PFTs, Table A1) using the standard CCI cross-walking scheme (CCI, 2017; Poulter et al., 2015) and aggregated to the common 0.25° grid. The annual PFT fractions are
 210 averaged over 2002–2020 to obtain a time-invariant representation of vegetation composition. These 11 fractional PFT layers, rather than categorical LC labels, are used as model predictors.

ESA CCI LC is also used for two evaluation purposes. First, for the tower-pixel representativeness analysis described in Sect. 3.1, tower and grid-cell LC is harmonised into three broad classes: forest, short vegetation, and cropland. Second, the 2016 ESA CCI LC map is aggregated by majority class into bare soil, sparse vegetation, grassland, cropland, shrubland,
 215 broadleaf deciduous forest (BDF), broadleaf evergreen forest (BEF), needleleaf evergreen forest (NEF), needleleaf deciduous forest (NDF), and mixed forest (MF), and used only to stratify and visualise GPP results.

2.1.5 Air temperature

We use T2M from ERA5-Land to train two GPP model configurations (Model A and Model B; Table 2) to compare our microwave-based LST models. T2M is a common proxy for the thermal control of autotrophic respiration in ecosystem carbon
 220 modelling (Atkin and Tjoelker, 2003), and ERA5-Land provides a consistent, widely used reanalysis source for this variable (Muñoz-Sabater et al., 2021).

2.1.6 Eddy-covariance data

We expand the FLUXNET2015 (Pastorello et al. (2020)) station ensemble used in Wild et al. (2022) with the FLUXNET Warm Winter (Warm Winter 2020 Team (2022)) and FLUXNET-CH₄ (Delwiche et al. (2021)) collections to extend tempo-
 225 ral coverage, add sites absent from FLUXNET2015, and broaden biome- and climate-space representation for training and validation.

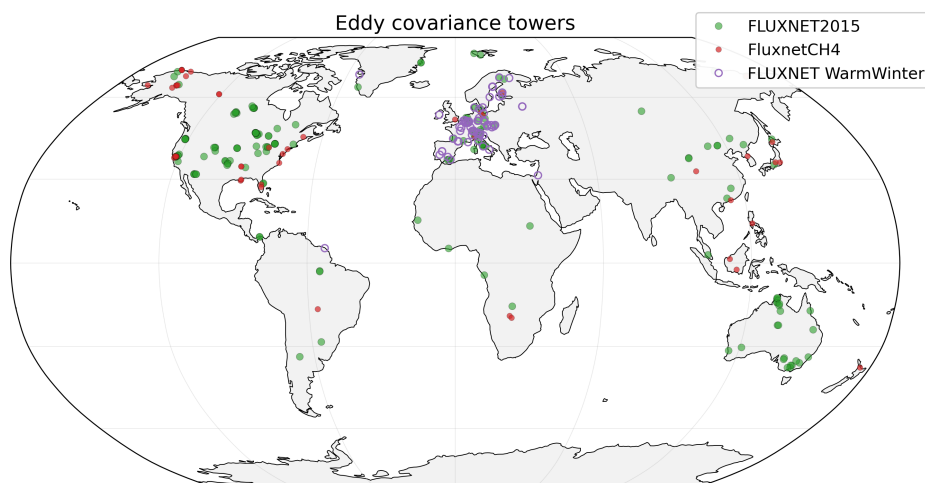


Figure 1. Spatial distribution of stations from FLUXNET2015, FLUXNET Warm Winter and FLUXNET-CH₄ datasets.

FLUXNET2015 provides harmonised EC time series of CO₂, water, and energy exchange together with meteorological and vegetation measurements from a global network of sites (Pastorello et al. (2020)). We use the February 2020 update, which increases the number of usable sites compared with earlier releases while retaining ONEFlux quality control and partitioning consistency across sites (Pastorello et al. (2020)). The FLUXNET Warm Winter (Warm Winter 2020 Team (2022)) release extends most European FLUXNET2015 sites to 2020 and adds stations, utilising a processing pipeline compliant with FLUXNET2015 to ensure that records are directly integrable (Warm Winter 2020 Team (2022)). FLUXNET-CH₄ compiles EC methane and CO₂ fluxes across a broad set of ecosystems, adding sites that are sparse or absent in FLUXNET2015, particularly wetlands and high-latitude systems (Delwiche et al. (2021)). Some towers appear in multiple collections, whereas others are unique to a single release Fig. 1. Therefore, we removed duplicates of sites available in multiple datasets, with a fixed priority order based on temporal completeness and continuity: *FLUXNET Warm Winter* > *FLUXNET2015* > *FLUXNET-CH₄*. Thus, if a site was present in Warm Winter, its Warm Winter record was retained as the source; otherwise, we selected FLUXNET2015, and used FLUXNET-CH₄ only when a site was absent from the other two. This policy reflects the generally longer temporal coverage of Warm Winter for European sites and the better consistency of FLUXNET2015 outside Europe, while FLUXNET-CH₄ primarily contributes additional high-latitude sites with shorter or less continuous CO₂ records. Cross-dataset consistency checks across LC and climate zones over overlapping periods (not shown) indicated high agreement between the three records, supporting the merged use of these sources as a single in situ GPP target ensemble. At each site, GPP was obtained from EC net ecosystem exchange (NEE) using both daytime and nighttime partitioning schemes; we then formed a single daily GPP by averaging the two, following recommendations to combine partitioning estimates for robust daily aggregates (Teubner et al., 2020; Wild et al., 2022). We filtered the data with the quality flag provided with the datasets and excluded low-quality days, retaining only days with at least 50 % of observations. After deduplication and quality filtering, the FLUXNET training dataset



comprised 215 sites, with an average observational record of approximately 4.7 years per site (median 3.3 years), based on the number of valid 8-day GPP observations.

2.2 Evaluation datasets

250 The models proposed in this manuscript are compared with external GPP products and complementary indicators of photosynthetic and hydrological variability. These datasets are used as evaluation rather than as error-free validation targets: they differ in observational inputs, model structure, training data, spatial and temporal resolution, and some share environmental drivers or EC observations with the present framework. The comparison, therefore, focuses on consistency of spatial and temporal behaviour across methodologically distinct products.

255 2.2.1 MODIS GPP

MODIS GPP builds on the light-use efficiency (LUE) framework of Monteith (1972), in which photosynthetic carbon uptake is proportional to the amount of photosynthetically active radiation absorbed by the canopy (APAR) and a biome-specific conversion efficiency. The operational MODIS algorithm (MOD17) implements this concept by combining MODIS-derived fraction of APAR (fAPAR) and ancillary radiation with ecophysiological scalars and biome parameters to estimate GPP at the 260 8 daily scale (Zhao et al., 2005; Running et al., 2015). In this study, we use the MOD17A2H Collection 6 product (8-daily at 500 m sampling, 2003-2020) and regrid it by averaging to $0.25^\circ \times 0.25^\circ$ to match the grid of our models for intercomparison and evaluation.

2.2.2 FLUXCOM-RS and FLUXCOM X-Base GPP

FLUXCOM provides global GPP fields by upscaling eddy-covariance fluxes with machine-learning ensembles (Tramontana 265 et al., 2016; Jung et al., 2020). We use FLUXCOM-RS (Jung et al., 2020), which is based solely on remotely sensed data, using MODIS-derived surface reflectance vegetation indices (NDVI, EVI, NDWI), fAPAR, and daytime and nighttime LST as predictors, thereby capturing the observed spatio-temporal variability of vegetation states without any meteorological input. For evaluating our models, we adopt FLUXCOM-RS to emphasise Earth observation-driven variability at fine spatial and 8-day temporal resolutions (native ~ 10 km). We additionally use FLUXCOM X-Base (X-BASE; Nelson & Walther et al. (2024)), the 270 successor upscaling framework to the original FLUXCOM ensemble, which combines remotely sensed vegetation indices and LST with meteorological drivers (actual and potential incoming shortwave radiation, air temperature, VPD) within a single, unified model, trained on an expanded eddy-covariance network of 294 sites at hourly, 0.05° resolution. FLUXCOM-RS for 2003-2015 is natively provided at 0.0833° (~ 10 km) and 8-day resolution, whereas FLUXCOM X-BASE is natively simulated at hourly and 0.05° resolution. FLUXCOM-RS (2003-2020) and FLUXCOM X-Base were aggregated to 0.25° to match the 275 grid of our models and to 8-daily where necessary.



2.2.3 DrylANNd GPP

As a dryland-specific benchmark, we use GPP from the DrylANNd (Dryland Artificial Neural Network) model (Dannenberg et al., 2023) for 2015–2020. DrylANNd estimates GPP, NEE, and evapotranspiration using an ensemble of feed-forward artificial neural networks trained on AmeriFlux eddy-covariance observations across arid to sub-humid ecosystems of the western United States. The model integrates multi-source satellite observations, combining optical vegetation indices (e.g., NDVI, NIRv, EVI), MODIS LST, and SMAP SM and soil temperature to better represent moisture limitation and pulse-driven productivity in drylands. Outputs are provided at monthly temporal resolution and 0.05° spatial resolution for the years 2015 – 2020.

We spatially average DrylANNd GPP to $0.25^\circ \times 0.25^\circ$ and use the native monthly estimates for intercomparison.

2.2.4 TRENDY ensemble v7 GPP

As an external process-based benchmark alongside the satellite-based products, we use GPP from the reanalysis-driven TRENDY-v7 ensemble of DGVMs (Le Quéré et al., 2018; Sitch et al., 2015) for the period 2003–2017. The v7 protocol prescribes historical variations in climate, land use, and atmospheric CO_2 to simulate GPP over 1950–2017. The ensemble comprises 16 DGVMs (CABLE-POP, CLASS-CTEM, CLM5.0, DLEM, ISAM, JSBACH, JULES, LPJ, LPJ-GUESS, LPX, OCN, ORCHIDEE, ORCHIDEE-CNP, SDGVM, SURFEX, VISIT), delivered at monthly time step on a common grid. We regridded each model to 0.25° using nearest-neighbour sampling and formed an unweighted ensemble-mean GPP time series for evaluation. For the temporal agreement analyses, which include multiple global GPP datasets, we interpolate TRENDY to 8-daily observations to enable direct comparison over the same spatial and temporal domain.

2.2.5 SPEI

To assess the response of predicted GPP to drought, we use the Standardised Precipitation–Evapotranspiration Index (SPEI; (Vicente-Serrano et al., 2010)), which quantifies anomalies in the climatic water balance (precipitation minus potential evapotranspiration, PET) relative to a long-term baseline. Negative values indicate dry conditions ($\text{SPEI} \leq -1$ moderate drought; $\text{SPEI} \leq -2$ extreme drought). We use the ERA5-based Global Drought Observatory implementation provided by the Joint Research Centre of the European Commission, where precipitation and PET are derived from ERA5 reanalysis (European Commission, Joint Research Centre (JRC), 2025a, b, c). Because SPEI integrates water supply and atmospheric demand, it has been widely used to diagnose drought-induced reductions in GPP and assess ecosystem sensitivity to water variability (Vicente-Serrano et al., 2013; Humphrey et al., 2018).

Here, we use SPEI for the period 2003–2020 with integration times of 1, 3, and 6 to capture short-term moisture-stress effects on GPP (SPEI-1, European Commission, Joint Research Centre (JRC) (2025a)) and more persistent seasonal water deficits (SPEI-3, SPEI-6, European Commission, Joint Research Centre (JRC) (2025b, c)). We therefore use SPEI as an external climatic indicator to assess whether the modelled GPP products capture drought-induced reductions in productivity. SPEI is provided at monthly and $0.25^\circ \times 0.25^\circ$ resolution, so resampling is not necessary.



2.2.6 TROPOSIF SIF

We use the TROPOSIF global sun-induced fluorescence (SIF) product derived from TROPOMI onboard Sentinel-5P (Guanter et al., 2021) to assess spatial and temporal variations in GPP. SIF is emitted as a by-product of photosynthesis and has been shown to closely track ecosystem GPP, often more directly than reflectance-based vegetation indices (Guanter et al., 2014; Zhang et al., 2016). TROPOSIF provides daily global coverage at high spatial resolution ($\sim 3.5 \times 5.5 \text{ km}^2$) and retrieves far-red SIF (743–758 nm) using a data-driven Fraunhofer line infilling approach. In this study, we employ the clear-sky L2B variant (cloud fraction ≤ 0.2) for the period 2018–2020 and aggregate the observations to a $0.25^\circ \times 0.25^\circ$ grid by averaging all valid retrievals within each target cell. Only measurements fulfilling the recommended quality filtering (qa_value > 0.5) and additional geometric constraints (solar zenith angle $\leq 70^\circ$, viewing zenith angle $\leq 60^\circ$) are retained to ensure robust clear-sky conditions.

2.3 Model implementations

Table 2. Overview of model configurations. Model A follows the GAM-based VODCA2GPP (Teubner et al., 2020; Wild et al., 2022) formulation, whereas Models B–I use RF regression. VOD denotes the combined use of three predictors: VOD, dVOD, and medVOD; LST denotes the joint use of both LST_{day} and LST_{night} ; SSM and RZSM denote the use of two predictors each: full signal and the corresponding anomaly.

Model	Algorithm	T2M	LST	VOD	Structural predictors	Hydrological predictors
A	GAM	×		×	–	–
B	RF	×		×	–	–
C	RF		×	×	–	–
D	RF		×	×	L-VOD	–
E	RF		×	×	LC	–
F	RF		×	×	L-VOD + LC	–
G	RF		×	×	L-VOD + LC	SSM
H	RF		×	×	L-VOD + LC	RZSM
I	RF		×	×	L-VOD + LC	SSM + RZSM

The model configurations provide a set of controlled contrasts to examine three aspects of the framework: the influence of the learning algorithm, the replacement of reanalysis T2M by microwave-derived LST, and the additional information provided by structural, LC, and hydrological predictors (Table 2). All configurations are compared using cross-validation and ALE, while selected models are retained for the subsequent global evaluation (Sect. 2.4).



Model A follows the original GAM-based VODCA2GPP formulation (Teubner et al., 2020; Wild et al., 2022), using T2M together with VOD, dVOD, and medVOD. Model B uses the same predictors in a Random Forest (RF), so that A–B isolates the effect of the learning algorithm. RFs are used for the remaining configurations because they can flexibly capture nonlinear relationships and interactions among heterogeneous predictors without a priori specification (Breiman, 2001; Tramontana et al., 2016; Schmidt et al., 2022). Model C replaces T2M with microwave-derived LST_{day} and LST_{night} , providing the microwave-driven thermal baseline.

Relative to Model C, Models D and E add L-VOD and LC fractions independently, while Model F combines both. Models G–I then evaluate the additional hydrological information from SSM, RZSM, and their combination, respectively. These configurations are treated as targeted contrasts rather than as a sequential forward-selection procedure because a predictor’s contribution depends on the information already available to the model.

Training samples are obtained by collocating the gridded predictors with EC GPP at each tower location. Towers within the same 0.25° grid cell share the predictor time series but retain their individual GPP observations. All configurations use the same common EC sample, and missing predictor values are not imputed. RF models are implemented in scikit-learn (Pedregosa et al., 2011), with hyperparameters selected through a systematic parameter search (Table A2). Configurations retained for the global analyses are subsequently trained on the full common EC sample and applied at 8-day and 0.25° resolution.

2.4 Model evaluation

Model evaluation proceeds from diagnostic assessment of the full set of models to targeted global comparisons. Models A–I are first evaluated at EC sites and through accumulated local effects (ALE) to assess predictive performance and the conditional response to individual predictors. Based on these diagnostics, Model H is retained as the extended RF configuration for the global evaluation. Models A, B, and H are compared for GPP magnitude to distinguish the effect of the learning algorithm (A–B) from that of the extended predictor framework (B–H). Subsequent analyses of training-network sensitivity and agreement with external datasets compare Models B and H, thereby retaining a common RF architecture.

2.4.1 Site-level performance and predictor responses

All model configurations are evaluated using site-grouped 10-fold cross-validation, with all observations from an EC site assigned entirely to either training or validation within each fold. Predictions for held-out sites are combined to obtain out-of-fold estimates for the complete validation network. Pearson’s correlation coefficient (r), Nash–Sutcliffe efficiency (NSE; Nash and Sutcliffe, 1970), RMSE and bias are calculated separately for each site at 8-day and monthly resolution and summarised by their median across the common set of validation sites. Performance is additionally stratified between forest and short-vegetation sites. To assess tower–pixel representativeness, we compare performance between sites where the broad tower LC class agrees with that of the corresponding 0.25° ESA CCI grid cell and sites where it does not (Sect. 2.1.4).

Conditional predictor responses are examined using ALE (Apley and Zhu, 2020), which estimates local changes in model output within the observed predictor distribution and is therefore well suited to correlated predictors (Schmidt et al., 2022; Kuhn-Régner et al., 2021). One-dimensional ALEs are calculated for Models A–I using a reproducible random sample of



up to 300,000 complete training observations per model, 20 quantile-based bins, and a minimum bin size of 50 observations. Effects outside the 5th–95th percentile range are interpreted cautiously because of lower data support. Two-dimensional ALEs for $LST_{\text{day}} \times \text{SSM}$ and $LST_{\text{day}} \times \text{RZSM}$ are additionally used to examine conditional water–temperature interactions.

2.4.2 GPP magnitude and training-network sensitivity

360 The magnitude and spatial distribution of GPP are evaluated for Models A, B, and H against MODIS, FLUXCOM-RS, FLUXCOM X-Base, and TRENDY-v7. Products are restricted to their common spatial domain ($115.1 \times 10^6 \text{ km}^2$) and the common 2003–2015 comparison period. Annual GPP is area-integrated over the valid domain; time steps with less than 50 % spatial coverage are excluded, and annual estimates are retained when temporal coverage exceeds 95 %. The remaining small temporal gaps are accounted for by annualisation. TRENDY is integrated on its native grid using fractional overlap with the common
 365 domain. Mean annual totals, latitudinal profiles, and LC-stratified distributions are used to identify both the magnitude and spatial origin of differences among products.

Sensitivity to the composition of the EC training network is evaluated for Models B and H by applying each of the ten cross-validation fold models globally (Tramontana et al., 2015). For each grid cell and year, the coefficient of variation (CV) across the ten-fold-specific annual GPP estimates is calculated as the standard deviation divided by the mean and expressed as a percentage; annual CV values are then averaged over the study period. Grid cells with mean annual GPP below $0.05 \text{ kg C m}^{-2} \text{ yr}^{-1}$
 370 are excluded because the CV becomes unstable near zero. This metric, therefore, quantifies sensitivity to training-site composition rather than total prediction uncertainty.

2.4.3 Global temporal agreement

Temporal agreement is evaluated for Models B and H against MODIS, FLUXCOM-RS, FLUXCOM X-Base, and TRENDY-
 375 v7. Model and reference GPP are collocated in space and time over their respective periods of overlap, with TRENDY monthly values interpolated to the common 8-day time axis for temporal alignment. Pixel-wise Pearson correlations are calculated for both the full GPP signal and 8-day anomalies. Anomalies are obtained by subtracting the corresponding 8-day-of-year climatology, thereby removing the mean seasonal cycle and emphasising sub-seasonal and interannual co-variability. Correlations are summarised by the LC classes defined in Sect. 2.1.4.

380 2.4.4 Evaluation of water-related GPP variability

Models B and H are additionally compared with DrylANNd GPP, TROPISIF SIF, and SPEI to evaluate GPP variability associated with photosynthetic activity and water limitation. For DrylANNd and SIF, model and reference observations are collocated in space and time, and Spearman rank correlations (ρ) are calculated pixel-wise for the full and anomaly signals; only grid cells with at least 24 valid paired observations are retained. Agreement is summarised by the median ρ and the fraction
 385 of negatively correlated grid cells. For SPEI, GPP anomalies are compared with SPEI at 1-, 3-, and 6-month integration times using Spearman correlation, with results stratified across the short-vegetation LC classes.



3 Results

3.1 Model performance

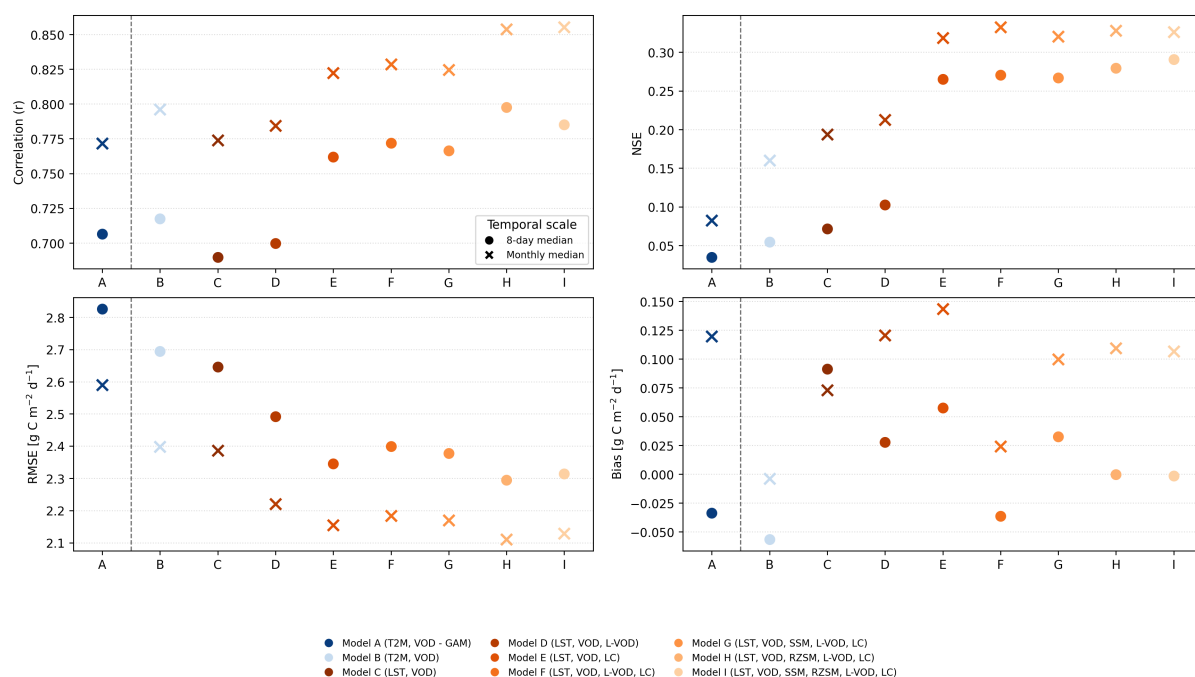


Figure 2. Median site-level cross-validation metrics computed over 10-folds for each model. Colors encode model family according to temperature input (blue: T2M; orange shades: LST). Panels report Pearson correlation (r), Nash–Sutcliffe efficiency (NSE), RMSE, and bias $\text{g C m}^{-2} \text{d}^{-1}$.

Site-grouped 10-fold cross-validation shows clear differences in median site-level performance across the model configurations (Fig. 2; Table A3).

Model B shows a modest improvement in predictive skill compared to Model A. At 8-day resolution, median correlation increases from approximately 0.71 to 0.72 and RMSE decreases from about $2.83 \text{ g C m}^{-2} \text{d}^{-1}$, while NSE also increases slightly. This indicates that, as expected, the choice of learning algorithm affects performance.

Replacing ERA5-Land T2M with microwave-derived LSTday and LSTnight (Model B vs. C) increases NSE and decreases RMSE further to approximately $2.65 \text{ g C m}^{-2} \text{d}^{-1}$. However, the correlation decreases slightly, consistent with the greater short-term variability and observational noise of satellite-derived LST relative to the smoother reanalysis T2M (Parinussa et al., 2011). Thus, microwave LST provides broadly comparable predictive information to T2M, with some improvement in error-based metrics but no systematic gain across all measures. This is consistent with previous flux-upscaling studies showing that remotely sensed LST contains useful information on ecosystem carbon flux variability (Tramontana et al., 2015, 2016; Jung et al., 2020).



Adding static L-VOD to LST+VOD (Model C → D) improves the model more consistently across metrics. This suggests that L-VOD provides complementary information not contained in the higher-frequency VOD predictors alone, consistent with its intended role as a structural constraint on vegetation state.

The greatest improvement is obtained by adding LC fractions to LST+VOD (Model C → E). Median 8-day r increases to approximately 0.76 and NSE to about 0.27, while RMSE decreases to roughly $2.35 \text{ g C m}^{-2} \text{ d}^{-1}$. Median bias remains relatively small and positive. The magnitude of this improvement indicates that explicitly representing differences among vegetation types is particularly important for transferring the GPP relationship across sites, consistent with the central role of biome and PFT information in flux-upscaling frameworks (Tramontana et al., 2015, 2016; Jung et al., 2020). Adding L-VOD together with LC fractions (Model E → F) produces only small and metric-dependent changes.

Adding SSM (Model G) does not produce a consistent improvement over Model F. Median r , NSE, and RMSE remain very similar, while bias shifts back towards a small positive value. In contrast, adding RZSM (Model H) yields a clearer gain: median 8-day r increases to about 0.80 and RMSE decreases to approximately $2.30 \text{ g C m}^{-2} \text{ d}^{-1}$. Model H also has a median bias close to zero, indicating that the gain in predictive skill is not achieved at the expense of a stronger systematic offset. The LC-stratified cross-validation results (Figure A1) indicate that the benefit of RZSM is concentrated in short vegetation sites: replacing SSM with RZSM (Model G → H) markedly improves correlation and RMSE for short vegetation, whereas the response for forest sites is much smaller and does not show a corresponding improvement in RMSE.

Sites with matching tower and satellite LC (Fig. A3) generally showed higher median correlations, lower RMSE, and bias closer to zero compared to sites with mismatched LC, indicating that footprint heterogeneity and tower–pixel representativeness contribute appreciably to model uncertainty. However, these sites were retained because excluding them would remove 83 towers (38% of the dataset, Fig. A2), and an additional sensitivity test (not shown) indicated lower overall model performance when training without them.

3.2 Relationship between GPP and predictors

The ALE analysis indicates that VOD, T2M, LST_{day} , and RZSM exhibit some of the largest ranges of modelled effects on GPP, as reflected by their comparatively large ALE magnitudes (Fig. 3). The temperature responses are physiologically plausible and broadly consistent across model configurations. Both T2M (Fig. 3c) and LST_{day} (Fig. 3d) show a strong increase in predicted GPP from cold conditions towards warmer temperatures, consistent with temperature limitation of photosynthesis under cool conditions (Chang et al., 2021; Lin et al., 2012). The response generally levels off or declines at the warmest temperatures, although the exact shape depends on the model: the GAM-based T2M response exhibits a more distinct optimum around 293–295 K, whereas the RF responses tend to increase over a broader part of the sampled temperature range. Similarly, LST_{day} increases strongly towards approximately 298–300 K before plateauing or declining slightly. LST_{night} (Fig. 3h) has a smaller ALE magnitude and shows a predominantly positive response from cold to warm nighttime conditions, which tends to level off at the warmer end of the distribution rather than displaying the same pronounced optimum as the daytime temperature predictors.



VOD (Fig. 3 a) exhibits a consistent, near-monotonic positive modelled association with GPP, and shows one of the largest
 435 absolute ALE magnitudes across predictors and models. This indicates that VOD provides substantial information for the
 fitted GPP models, consistent with its sensitivity to vegetation water content and biomass. The other VOD-derived predictors
 provide complementary information: *d*VOD (Fig. 3g) generally shows increasing GPP effects with positive vegetation changes,
 while medVOD (Fig. 3b) has a predominantly negative conditional response that weakens once L-VOD and LC fractions
 are included. L-VOD (Fig. 3e) shows a comparatively stronger response when introduced without LC fractions (Model D),
 440 but a much smaller independent effect in Models F–I, suggesting substantial overlap between L-VOD and LC fractions in
 representing persistent vegetation structure.

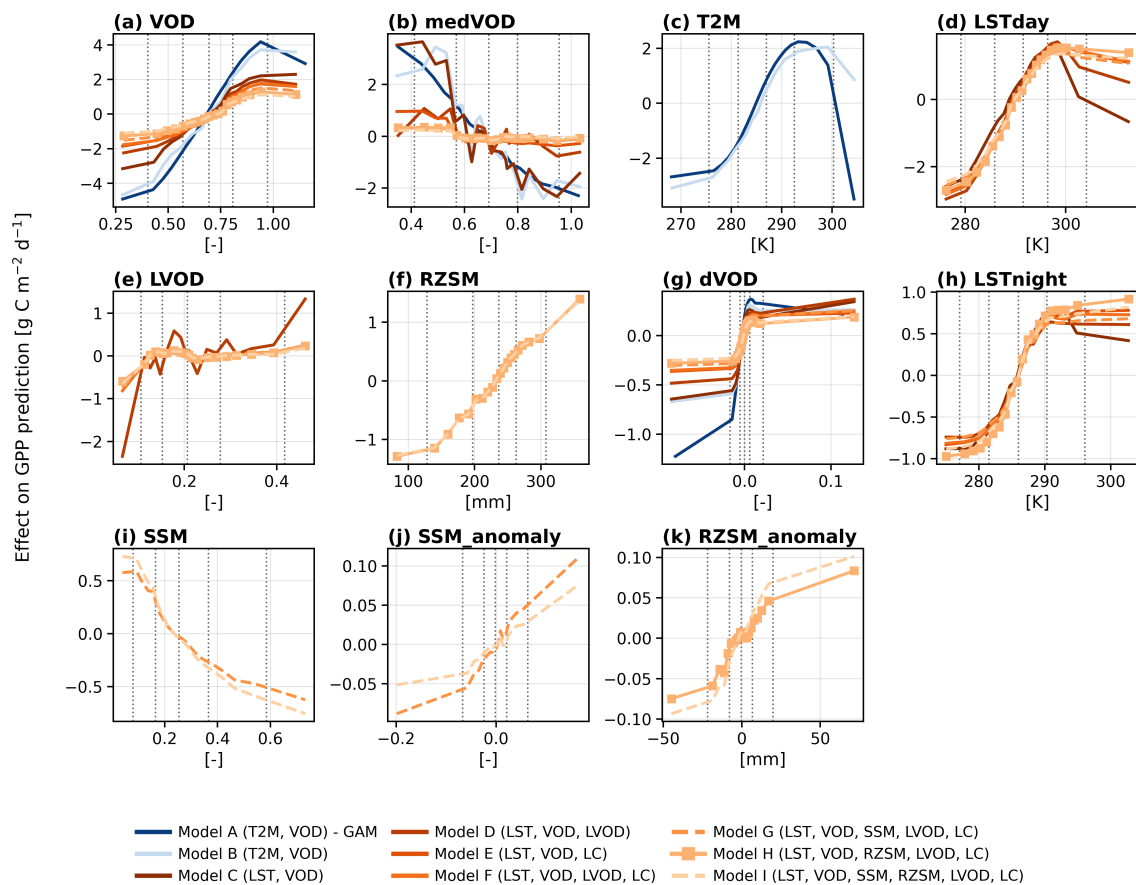


Figure 3. ALEs of non-PFT predictors across all models. Curves are centred 1D ALE (20 quantile bins) with dotted vertical lines at the 0.05/0.25/0.50/0.75/0.95 quantiles of predictor value. Positive/negative ALE values indicate model effects above/below the centred mean effect over the data used for the ALE calculation.



RZSM (Fig. 3 f) rises approximately monotonically, a modelled relationship consistent with increasing water availability alleviating stomatal limitation.

SSM (Fig. 3i) shows a weak negative modelled effect, which may partly reflect its role as a short-term weather proxy: wetter surface soils often follow precipitation events that coincide with lower radiation and cooler conditions (Koirala et al., 2017; Walther et al., 2019). Similar negative GPP–precipitation sensitivities have been reported in some ecosystems and seasons, including energy-limited sites (Miller et al., 2023). The 2D ALEs provide additional evidence for interactions between water and thermal conditions (Fig. A5). The $SSM \times LST_{day}$ surface (Fig. A5) shows mostly weak interactions, with localised negative interaction effects at high SSM and sub-optimal LST_{day} . The $RZSM \times LST_{day}$ surface (Fig. A5) shows a stronger positive interaction around intermediate-to-warm temperatures (≈ 290 – 300 K), indicating that the joint model response under these conditions exceeds that expected from the additive RZSM and LST_{day} effects alone. This pattern is consistent with coupled water-energy constraints on GPP.

In terms of LC fractions (Fig. A4), the ALEs show that increasing tree cover has the largest positive effect on predicted GPP, especially for evergreen (pftTreeBE; Fig. A4 a) and deciduous (pftTreeBD; Fig. A4 b) broadleaf and needleleaf evergreen (pftTreeNE; Fig. A4 c) trees, with gains that rise steeply at low-moderate cover and then saturate at higher fractions. Shrub cover (pftShrub* panels; Figs. A4 c, d, e) exhibits a negative effect, strongest for broadleaf shrubs, likely reflecting a compositional trade-off: increasing shrub cover typically comes at the expense of tree cover, leading to lower productivity. Bare fraction (pftBare; Fig. A4 d) is also negative, as expected for low-productivity surfaces. The near overlap of curves across moisture-augmented models indicates that adding SM/RZSM primarily adjusts magnitude rather than sign or shape, reinforcing that LC fractions provides a stable structure for the RF to separate productive (tree-dominated) from low-productivity (shrub/bare) landscapes.

3.3 Magnitude of predicted GPP

Models A, B, and H differ substantially in their integrated GPP magnitude (Table A4). Model A produces $206.8 \text{ Pg C yr}^{-1}$, a higher magnitude than Model B ($189.4 \text{ Pg C yr}^{-1}$). This indicates that the choice of algorithm alone lowers the estimated GPP magnitude. Adding additional predictors in Model H yields a further reduction ($175.0 \text{ Pg C yr}^{-1}$). Despite its lower GPP magnitude, Model H remains substantially higher than the reference products over the common spatial domain. MODIS, FLUXCOM-RS, and FLUXCOM X-Base estimate 111.1 , 108.4 , and $121.5 \text{ Pg C yr}^{-1}$, respectively, while individual TRENDY-v7 members span 72.1 – $148.0 \text{ Pg C yr}^{-1}$.

3.3.1 Latitudinal distribution of predicted GPP

The latitudinal profiles in Fig. 4 show the spatial origin of the differences in integrated GPP. All products reproduce the expected zonal pattern of high tropical productivity and declining GPP towards the subtropics and high latitudes, although their magnitudes differ substantially (Beer et al., 2010; Anav et al., 2015).

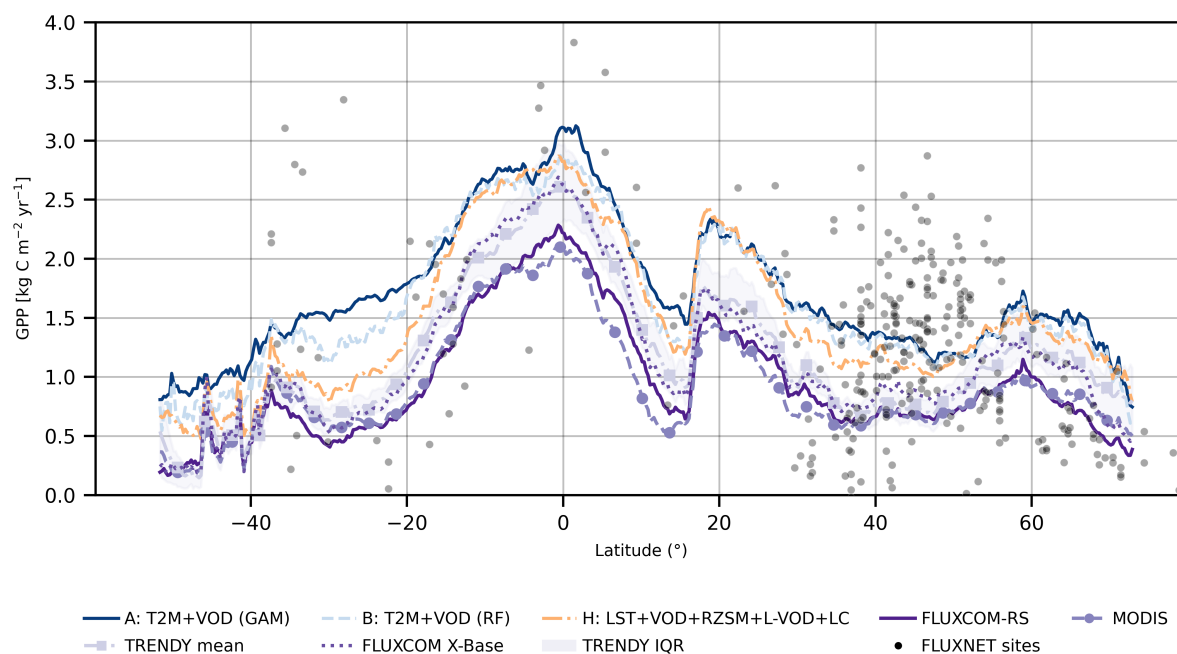


Figure 4. Latitudinal mean GPP for model configurations A, B, H and reference products. Means are calculated from observations collocated in space and time over the common comparison period. The TRENDY-v7 distribution is shown as the ensemble mean and interquartile range.

Differences among Models A, B, and H vary substantially with latitude. Replacing the GAM formulation of Model A with RF (Model B) generally reduces GPP, while the largest difference between Models B and H occurs between approximately -35° and -20° latitude. In these broadly water-limited subtropical regions, Model H shows a pronounced reduction in GPP, bringing the estimates closer to the reference products.

At the tropical maximum, Models B and H fall within the TRENDY interquartile range and are also close to the FLUXCOM-X estimates, whereas Model A shows higher GPP. MODIS and FLUXCOM-RS show substantially lower estimates than our configurations; however, MOD17 has been reported to underestimate at highly productive sites (Turner et al., 2006).

At high northern latitudes, a substantial discrepancy also persists. Above approximately 70°N , Models A, B, and H remain generally higher than MODIS, FLUXCOM-RS, and FLUXCOM X-Base, which converge towards $\approx 0.3\text{--}0.7 \text{ kg C m}^{-2} \text{ yr}^{-1}$, although the model estimates mostly remain within the TRENDY interquartile range. The particularly large departures of Model A in the Southern Hemisphere extratropics and at high northern latitudes coincide with regions that are relatively poorly represented by FLUXNET.

3.3.2 Biome dependence of GPP magnitude.

The biome distributions in Fig. 5 further identify the ecosystem types underlying the differences in GPP magnitude. Across all products, the broad ecological ordering is consistent: BEF has the highest productivity, whereas shrublands and sparse veg-



etation occupy the low-productivity end of the distribution, reflecting differences in growing-season length, canopy structure, water availability, and absorbed radiation (Beer et al., 2010; Anav et al., 2015).

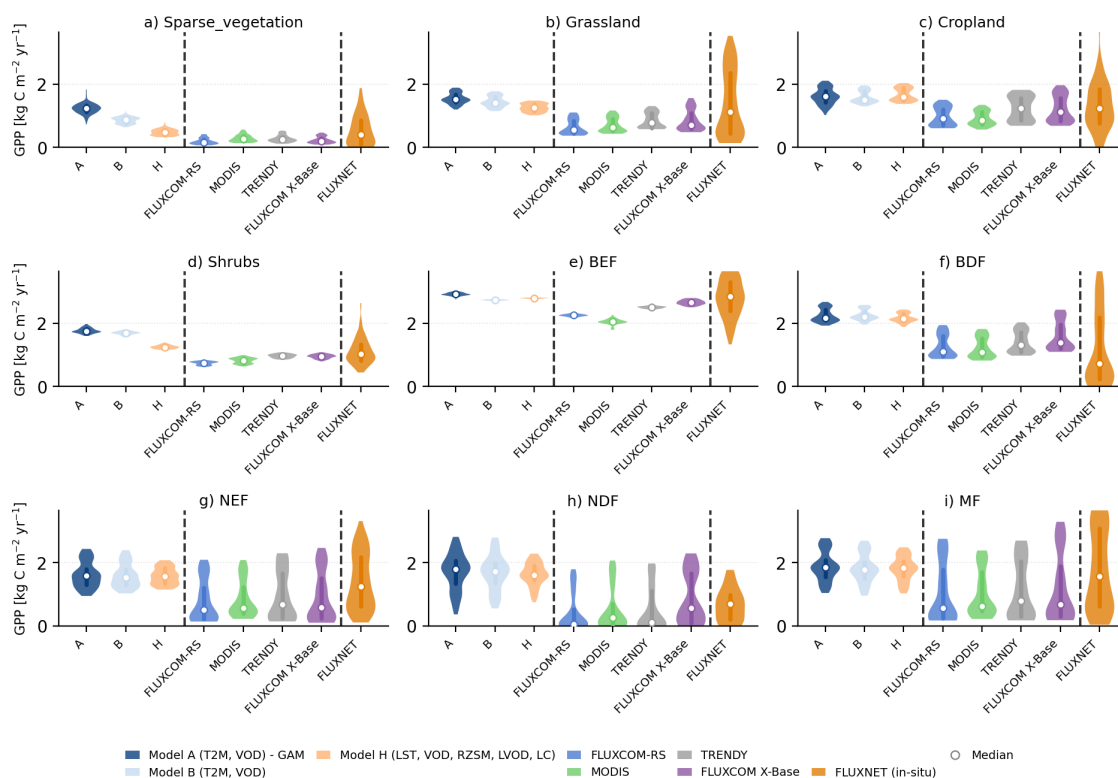


Figure 5. Distribution of GPP by land-cover class for Models A, B, H, FLUXNET observations, and the gridded reference products. Only observations collocated in space and time across the products are used. The distributions, therefore, diagnose biome-dependent differences in GPP magnitude independently of unequal sampling in the comparison datasets.

One of the largest differences between Models A, B, and H and the references occur in BEF, contributing to the tropical offset visible in Fig. 4. Model H shows substantially lower GPP than Models A and B in short vegetation biomes, bringing the estimates closer to the reference products and suggesting a better-constrained representation of productivity in water-limited ecosystems.

3.4 Sampling uncertainty

Global mean inter-fold CV decreases (Fig., 3.4, Table A5) from 14.2% in Model B to 7.9% in Model H, corresponding to a 44% reduction in sensitivity to the composition of the flux-tower training network. The largest uncertainties occur in sparsely sampled dry regions. In the Sahara/Sahel, mean CV decreases from 21.1% in Model B to 14.7% in Model H, while over arid Australia it decreases more modestly from 24.8% to 20.3%. Thus, although Model H substantially reduces fold sensitivity



globally, arid Australia remains a region of comparatively high uncertainty. A similarly strong reduction occurs in the boreal
 500 zone ($> 50^{\circ}\text{N}$), where mean CV decreases from 10.4% to 4.8%. Regions of persistently elevated CV broadly coincide with
 areas of sparse FLUXNET coverage.

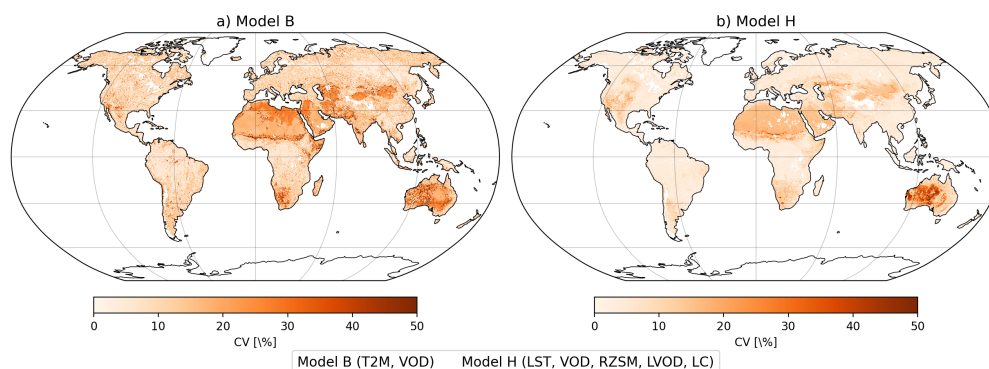


Figure 6. Inter-fold coefficient of variation of predicted GPP as a measure of uncertainty.

3.5 Temporal agreement with external GPP datasets

We evaluated Models B and H against four external GPP products (MODIS, FLUXCOM-RS, FLUXCOM X-Base, and TRENDY; Figs. 7 and A7, Table A8). For the full 8-day signal, Model H generally shows higher correlations than Model B
 505 across LC classes and reference products. Agreement is strongest in BDF, NEF, NDF, and MF, where median r for Model H ranges from approximately 0.76 to 0.88, while clear improvements are also evident in sparse vegetation, grasslands, croplands, and shrubs. BEF remains the main exception, with low and reference-dependent correlations, consistent with its weak season-
 ality and limited temporal dynamic range (Restrepo-Coupe et al., 2013; Guan et al., 2015; Tramontana et al., 2016; Wild et al., 2022).

510 The improvement from Model B to Model H is even clearer for 8-day anomalies (Fig. A7 and Table A8). Across sparse vegetation, grasslands, croplands, and shrubs, median correlations increase by approximately 0.11–0.29 across the four reference products. Improvements are also evident in BDF, NEF, NDF, and MF, although agreement with TRENDY is generally lower than with the observation-based products. Agreement in BEF remains comparatively weak and reference-dependent.

Overall, the comparison with four external GPP products supports Model H over the simpler Model B, particularly for
 515 reproducing short-term GPP variability across contrasting LC classes.

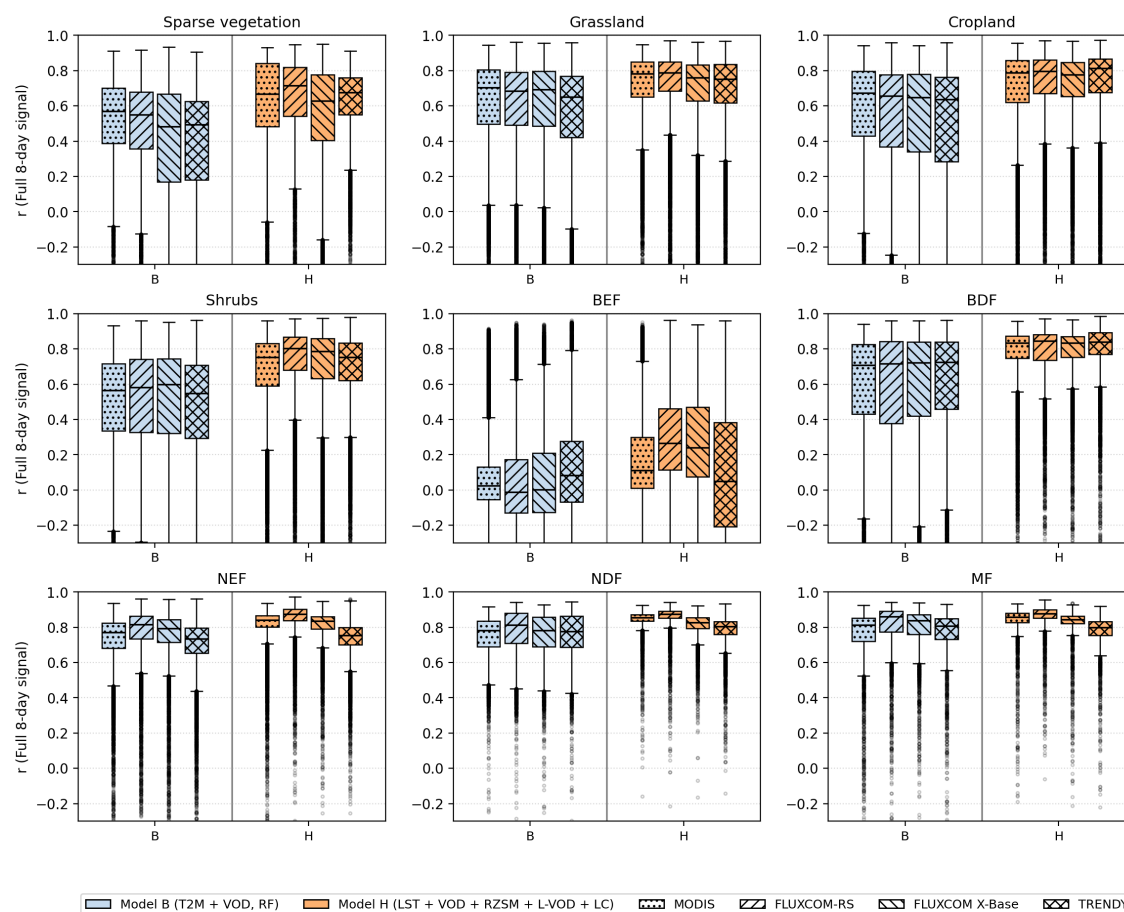


Figure 7. Pixel-wise Pearson correlation coefficient (r) between predicted GPP and FLUXCOM-RS, FLUXCOM X-Base, MODIS, and TRENDY on harmonised 8-day time series. Each model–reference pair is restricted to its shared temporal overlap, and boxplots are shown by LC class.

3.6 Evaluation of photosynthetic and water-related GPP variability using SIF, SPEI and DrylANNd GPP

Agreement with DrylANNd improves markedly from Model B to Model H (Fig. 8; Table A7). Median full-signal Spearman correlation increases from $\rho = 0.780$ to 0.841, while anomaly correlation increases from 0.183 to 0.452. The fraction of grid cells with negative anomaly correlations decreases from 16.6 % to 2.8 %.

520 SIF shows a similar, although weaker, improvement. Median full-signal correlation increases from $\rho = 0.422$ for Model B to 0.557 for Model H, and anomaly correlation increases from 0.064 to 0.124. The fraction of negative anomaly correlations decreases from 35.6 % to 27.0 %. Agreement remains comparatively weak for SIF anomalies, potentially reflecting the variable coupling between SIF and GPP under high atmospheric demand and drought (Li and Xiao, 2022; Qiu et al., 2024).

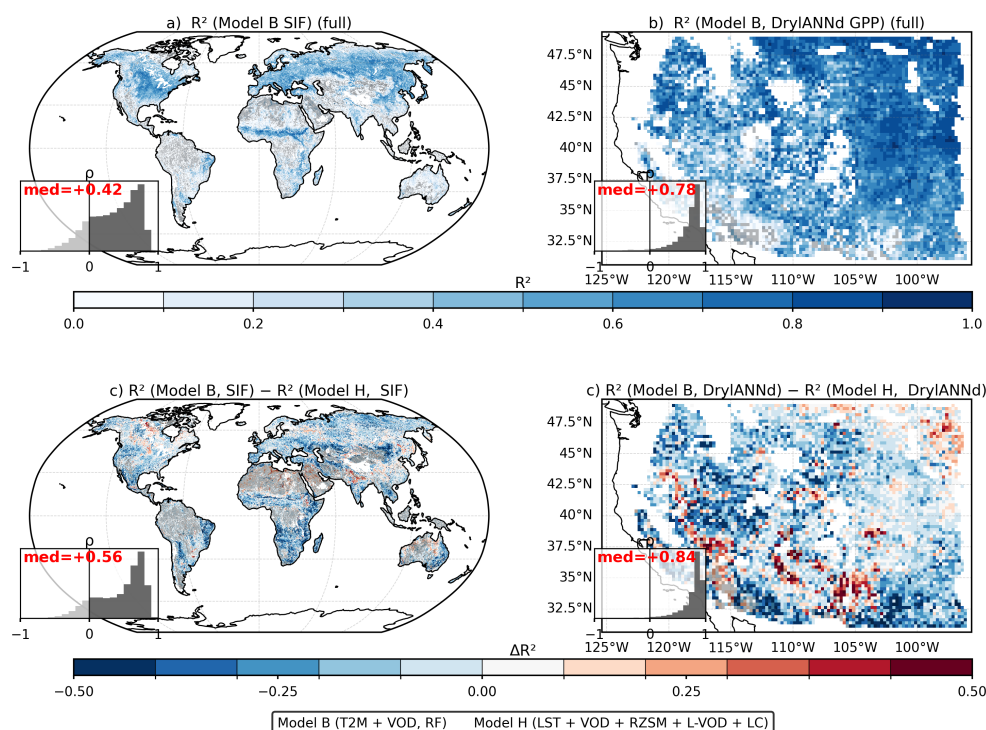


Figure 8. Spatial agreement over the full signal of Models B and H with SIF (left) and DrylANNd GPP (right). The upper panels show R^2 for Model B, while the lower panels show the difference between Models B and H. Blue indicates higher R^2 for Model H and red higher R^2 for Model B. Grey marks cells with negative Spearman ρ or fewer than 24 valid collocated observations. Insets show the distribution of Spearman ρ and its median.

The SPEI comparison likewise shows stronger coupling of Model H GPP anomalies to drought variability (Fig. 9). The largest differences occur in sparse vegetation and shrubs. For sparse vegetation, ρ increases from 0.41 to 0.75 for SPEI-03 and from 0.46 to 0.75 for SPEI-06; for shrubs, it increases from 0.30 to 0.62 and from 0.26 to 0.55, respectively. Grasslands also show a pronounced increase, from -0.02 to 0.26 for SPEI-03 and from -0.08 to 0.28 for SPEI-06.

Overall, Model H shows improved agreement with both photosynthetic variability represented by SIF and water-related GPP variability represented by DrylANNd and SPEI. The strongest improvements in the latter occur in short-vegetation classes, supporting an improved representation of water-related GPP dynamics relative to Model B.

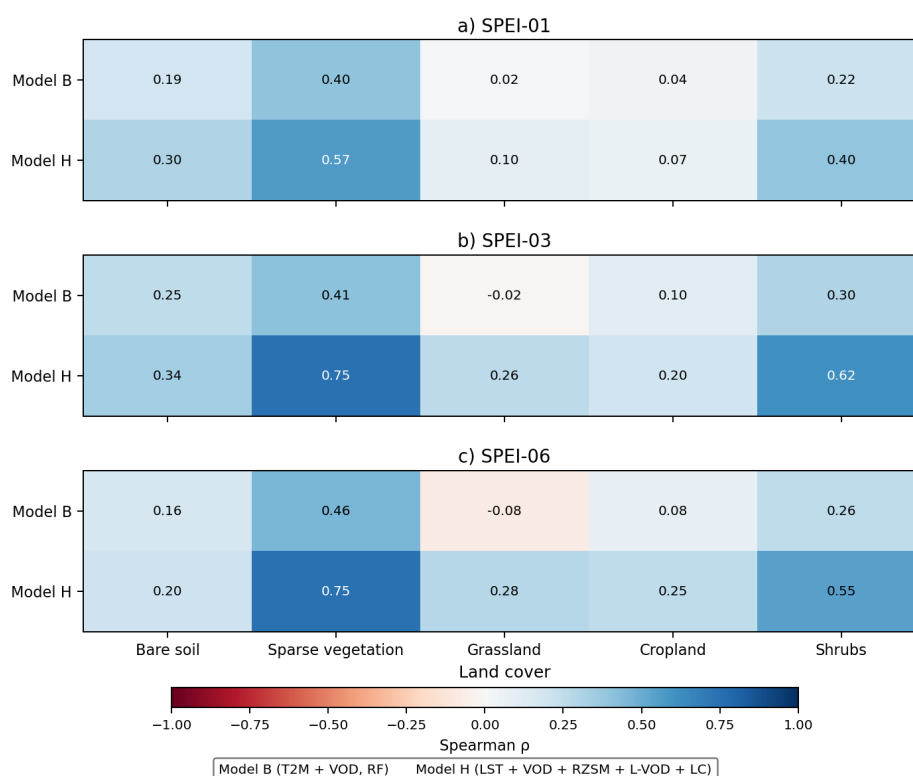


Figure 9. Spearman rank correlation (ρ) between GPP anomalies and SPEI at 1-, 3-, and 6-month timescales (SPEI-01, SPEI-03, SPEI-06) for five short-vegetation land-cover classes.

4 Discussion and conclusions

4.1 Feasibility of microwave-driven GPP estimation

The central question of this study is whether the main time-varying controls on GPP can be represented from microwave observations while retaining transferability across ecosystems. Previous studies established the potential of VOD for GPP estimation, but also showed that the VOD–GPP relationship depends on vegetation type and microwave frequency (Teubner et al., 2018, 2019, 2020; Wild et al., 2022). Our results show that this framework can be extended with microwave-derived thermal and hydrological information, although the individual predictors contribute differently.

Replacing ERA5-Land T2M with microwave-derived LST yields mixed but broadly comparable site-level performance. Error-based metrics improve slightly, whereas correlation decreases somewhat. This is consistent with radiometric LST sampling the instantaneous surface thermal state and exhibiting greater short-term variability than the smoother reanalysis T2M (Parinussa et al., 2011; Good et al., 2017). The comparison therefore supports LST as an observation-based thermal constraint,



rather than indicating a general increase in predictive skill. Importantly, it removes reanalysis temperature as a time-varying model input.

Canopy structure and vegetation composition have a strong effect on transferability. LC fraction provides the largest individual improvement among the tested predictor groups. Its 11 fractional PFT layers explicitly represent sub-grid vegetation composition, allowing otherwise similar microwave signals to be interpreted differently among ecosystems. L-VOD provides complementary information on persistent canopy structure, but its conditional effect becomes much smaller once LC is included, as also indicated by the ALE analysis. This suggests partial redundancy between these structural descriptors rather than a simple additive contribution of each predictor.

The hydrological comparisons further distinguish shallow from deeper soil-water information. Adding SSM does not consistently improve cross-validation performance, whereas RZSM produces a clearer gain, particularly at short-vegetation sites. Its approximately monotonic positive ALE response is also consistent with increasing root-zone water availability relieving water limitation. This behaviour is physically plausible because deeper SM dynamics integrate water availability over longer timescales than the rapidly varying surface layer; previous analyses likewise show stronger relationships between deeper SM and vegetation productivity in many grasses, shrubs, and semi-arid ecosystems (Li et al., 2021a). Including SSM together with RZSM provides no consistent additional benefit. The weak negative conditional SSM response should not be interpreted causally: surface wetness covaries with precipitation, cloudiness, radiation, and temperature and can therefore act as a short-term weather proxy (Koirala et al., 2017; Walther et al., 2019).

Because LC remains a static non-microwave predictor, the resulting framework is best described as microwave-driven rather than microwave-only. Its time-varying vegetation, thermal, and hydrological information is derived from microwave observations, while LC fractions provide the ecological context needed for transfer across vegetation types.

4.2 Global consistency and photosynthetic and water-related GPP variability

The comparisons with external GPP datasets indicate that the extended framework improves the representation of GPP dynamics beyond the EC sites. Model H generally agrees more closely than Model B with MODIS, FLUXCOM-RS, FLUXCOM X-Base, and TRENDY across LC classes, with particularly clear improvements for GPP anomalies in short vegetation. These comparisons evaluate the net effect of the extended Model H configuration and cannot attribute the improvement to RZSM, LC, or L-VOD individually; those contributions are instead diagnosed by the A–I cross-validation and ALE contrasts.

The complementary evaluations provide evidence for both photosynthetic and water-related variability. Agreement with SIF improves more modestly, particularly for anomalies, indicating broader consistency with an observation-based proxy of photosynthetic activity while also reflecting the variable coupling between SIF and GPP under drought and across aridity gradients (Li and Xiao, 2022; Qiu et al., 2024). In contrast, DrylANNd and SPEI provide more direct evidence for water-related GPP dynamics. Agreement with DrylANNd improves substantially for both the full signal and anomalies, consistent with better representation of productivity variability in drylands, where soil-water availability exerts a strong control (Liu et al., 2020; Kannenberg et al., 2024). Model H GPP anomalies also show stronger relationships with SPEI, especially at 3- and 6-month timescales in sparse vegetation, shrublands, and grasslands, indicating improved representation of persistent



water-related variability (Wang et al., 2022; Zhang et al., 2021; Vicente-Serrano et al., 2013). Together with the site-level RZSM contrasts, these results address an important weakness of the original VODCA2GPP implementation, which showed comparatively weak GPP responses with increasing drought severity (Wang et al., 2022).

Agreement with external products should nevertheless not be interpreted as validation against independent truth. Several products share EC observations, meteorological forcing, or related Earth-observation inputs with the present framework, and TRENDY is available only monthly. The comparisons are therefore most useful for identifying patterns that are robust across methodologically different estimates and ecosystems where disagreement persists.

BEF remains a notable example of such disagreement. Temporal correlations are weak and reference-dependent, while estimates of GPP magnitude also diverge substantially. Tropical productivity remains comparatively difficult to constrain because persistent cloud cover and saturation affect optical products and tropical EC sites remain sparse (Turner et al., 2006; Tramontana et al., 2016; Wild et al., 2022). The spread in BEF should therefore be interpreted as structural uncertainty among approaches rather than as unambiguous evidence of bias in any single product.

4.3 Absolute GPP magnitude and spatial extrapolation

Absolute GPP magnitude remains the main unresolved uncertainty. Model H produces lower GPP than Models A and B but remains substantially higher than MODIS, FLUXCOM-RS, FLUXCOM X-Base, and the TRENDY ensemble over their common domain. Such differences cannot be interpreted simply as positive bias, because the absolute magnitude of terrestrial GPP is itself poorly constrained and substantial spread exists among observation-based and process-based estimates (Beer et al., 2010; Anav et al., 2015; Jung et al., 2020). Recent atmospheric constraints also suggest that commonly used estimates may be too low. Cuntz et al. (2025), for example, estimated northern extratropical GPP ($> 23^{\circ}\text{N}$) at approximately $51 \pm 2 \text{ Pg C yr}^{-1}$, corresponding to a global value of about $151 \pm 4 \text{ Pg C yr}^{-1}$ in their scaling experiments. These ranges describe spread among the matching scenarios rather than formal uncertainty. Although closer to Model H than many of the external products considered here, those estimates are not directly comparable because the spatial domains and constraints differ.

The contrast between Models A and B further demonstrates that GPP magnitude depends strongly on model formulation. They use the same T2M+VOD predictors and EC target, yet the RF substantially reduces integrated GPP and the large latitudinal departures produced by the GAM, particularly in the Southern Hemisphere extratropics and at high northern latitudes. The further reduction in Model H indicates that both the learning algorithm and additional predictor constraints affect large-scale extrapolation. Consequently, the remaining magnitude discrepancy cannot be attributed solely to the EC target or to one missing predictor.

Spatial representativeness provides an additional source of uncertainty. EC measurements sample a much smaller and dynamically varying source area than the 0.25° satellite grid cell, and sites with mismatching tower and grid-cell LC show lower cross-validation performance (Chu et al., 2021; Fang et al., 2024). Uneven environmental and geographical coverage of the tower network further affects extrapolation. The fold-based analysis supports this interpretation: regions with sparse EC coverage, including arid Australia and parts of the Sahara and Sahel, remain particularly sensitive to training-site composition, although this sensitivity is lower for Model H than for Model B (Tramontana et al., 2015; Roberts et al., 2017; Sulkava et al.,



2011). This reduction reflects the net effect of the extended Model H configuration and cannot be assigned to LC, L-VOD, or RZSM individually.

Strong hydroclimatic variability in semi-arid ecosystems may amplify differences among relationships learned from different training subsets (Ahlström et al., 2015; Poulter et al., 2014), but it is important to distinguish this from the quantity measured here: the fold CV quantifies sensitivity to training-site composition, not climatic variability or total prediction uncertainty.

Reducing tower–pixel mismatch should therefore be a priority for future development. Footprint-aware matching and higher-resolution predictor data could improve the correspondence between EC observations and satellite inputs. Using AMSR-E and AMSR2 observations closer to their native resolution and future higher-resolution microwave observations from missions such as CIMR may further reduce sub-pixel mixing (Donlon et al., 2023).

4.4 Remaining environmental and observational constraints

The present predictor set does not explicitly describe all processes that can down-regulate photosynthesis. Incoming radiation and photoperiod are particularly relevant at high latitudes and during seasonal transitions, where similar vegetation, temperature, and soil-water conditions can occur under very different light availability. Atmospheric VPD is another missing constraint because high evaporative demand can reduce stomatal conductance even when root-zone water remains available (Nemani et al., 2003; Grossiord et al., 2020). These missing controls may contribute both to regional differences and to uncertainty in absolute GPP magnitude.

The EC target also introduces uncertainty. GPP is not measured directly but derived from NEE partitioning, and daytime- and nighttime-based approaches can produce systematically different estimates (Reichstein et al., 2005; Lasslop et al., 2010; Pastorello et al., 2020). Because all configurations are trained against the same averaged target, partitioning uncertainty cannot explain the large differences among Models A, B, and H, but it can influence their common absolute scale.

A further limitation arises from the multi-sensor LST record. AMSR-E/AMSR2 and SSMIS F17 observe at different local times, and the DMSP orbit is subject to drift (Yang et al., 2011). CDF harmonisation reduces distributional offsets among sensors but cannot remove differences caused by sampling different phases of the diurnal temperature cycle. Residual observation-time and orbital effects therefore remain a source of uncertainty in the LST record and consequently in GPP.

4.5 Synthesis and outlook

Taken together, the results support the feasibility of microwave-driven GPP upscaling while defining where the present framework remains uncertain. Microwave-derived LST provides a broadly comparable alternative to reanalysis T2M, LC supplies the strongest ecological context for transferring microwave signals among ecosystems, L-VOD adds complementary canopy structure information, and RZSM improves the representation of water-related variability, particularly at short-vegetation sites. Model H consequently provides a useful balance between site-level performance and representation of hydrologically driven GPP dynamics.

The principal remaining challenges concern absolute GPP magnitude, spatial extrapolation, and environmental controls not explicitly represented by the current predictors. Future development should combine footprint-aware training and higher-

resolution microwave observations with additional constraints on radiation and atmospheric water demand. Active-microwave observations may also provide complementary structural information; for example, ESA's BIOMASS mission uses P-band SAR to retrieve forest above-ground biomass and height (Quegan et al., 2019). Combining such structural information with the long-term passive-microwave record could improve spatial transferability while preserving the observational complementarity that motivates a microwave-driven GPP product.

Data availability. Data from Model H will be made available after review in an open TU Wien Research Data <https://researchdata.tuwien.ac.at> repository.

Appendix A: Appendix heading

Table A1. 11 Plant Functional Type (PFT) Predictors used in this study

Predictor	Vegetation			
	Leaf type	Phenology	Growth Form	Other
pftTreeBE*	Broadleaved	evergreen	Tree	–
pftTreeBD*	Broadleaved	deciduous	Tree	–
pftTreeNE*	Needle-leaved	evergreen	Tree	–
pftTreeND*	Needle-leaved	deciduous	Tree	–
pftShrubBE*	Broadleaved	evergreen	Shrub	–
pftShrubBD*	Broadleaved	deciduous	Shrub	–
pftShrubNE*	Needle-leaved	evergreen	Shrub	–
pftHerb [†]	–	–	Herbaceous cover	–
pftCrop [†]	–	–	Cropland	–
pftBare*	–	–	–	Bare soil
pftNoLand*	–	–	–	No land



Table A2. Random forest hyperparameter search space and selected values. Detailed descriptions of the hyperparameters are provided in the scikit-learn documentation (Pedregosa et al., 2011).

Hyperparameter	Search range	Selected value	Description
n_estimators	{200, 400, ..., 2000} (step 200)	1200	Number of trees in the forest.
max_features	{1, 2, ..., p } or {sqrt, log2}	5	Number of features considered at each split.
min_samples_split	{2, 5, 10, 15, 20}	5	Minimum number of samples required to split an internal node.
min_samples_leaf	{1, 2, 4, 8, 15, 20}	15	Minimum number of samples required at a leaf node.
bootstrap	{True, False}	True	Whether bootstrap samples are used when building trees.

Table A3. Cross-validation performance at 8-day (8D) and monthly (M) temporal scales. Values represent median out-of-fold site-level metrics calculated across the common set of 215 validation sites.

Model	Scale	r	NSE	RMSE	Bias
Model A (T2M, VOD) – GAM	8D	0.707	0.035	2.826	-0.034
Model A (T2M, VOD) – GAM	M	0.772	0.082	2.590	0.120
Model B (T2M, VOD)	8D	0.718	0.055	2.695	-0.057
Model B (T2M, VOD)	M	0.796	0.160	2.398	-0.004
Model C (LST, VOD)	8D	0.690	0.072	2.646	0.091
Model C (LST, VOD)	M	0.774	0.194	2.386	0.073
Model D (LST, VOD, L-VOD)	8D	0.700	0.102	2.491	0.028
Model D (LST, VOD, L-VOD)	M	0.784	0.213	2.220	0.120
Model E (LST, VOD, LC)	8D	0.762	0.265	2.345	0.058
Model E (LST, VOD, LC)	M	0.822	0.319	2.154	0.143
Model F (LST, VOD, L-VOD, LC)	8D	0.772	0.271	2.399	-0.036
Model F (LST, VOD, L-VOD, LC)	M	0.828	0.333	2.183	0.024
Model G (LST, VOD, SSM, L-VOD, LC)	8D	0.766	0.267	2.378	0.033
Model G (LST, VOD, SSM, L-VOD, LC)	M	0.825	0.320	2.169	0.100
Model H (LST, VOD, RZSM, L-VOD, LC)	8D	0.798	0.279	2.294	0.000
Model H (LST, VOD, RZSM, L-VOD, LC)	M	0.854	0.328	2.111	0.109
Model I (LST, VOD, SSM, RZSM, L-VOD, LC)	8D	0.785	0.291	2.314	-0.001
Model I (LST, VOD, SSM, RZSM, L-VOD, LC)	M	0.855	0.326	2.129	0.107

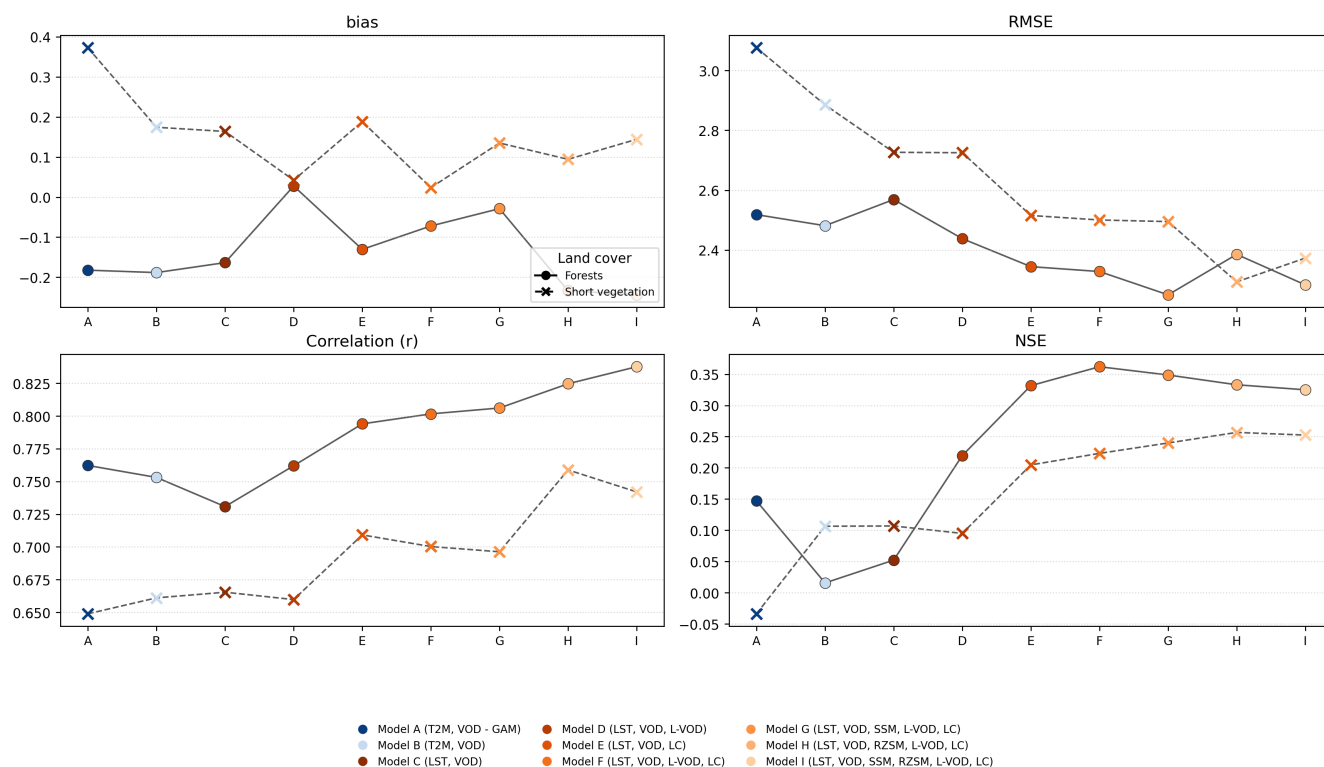


Figure A1. Median of per-station cross-validation metrics, by model and land-cover group. Forests = BEF, BDF, NDF, NEF and MF forests; short vegetation = shrubs, sparse vegetation, grassland, and bare soil.

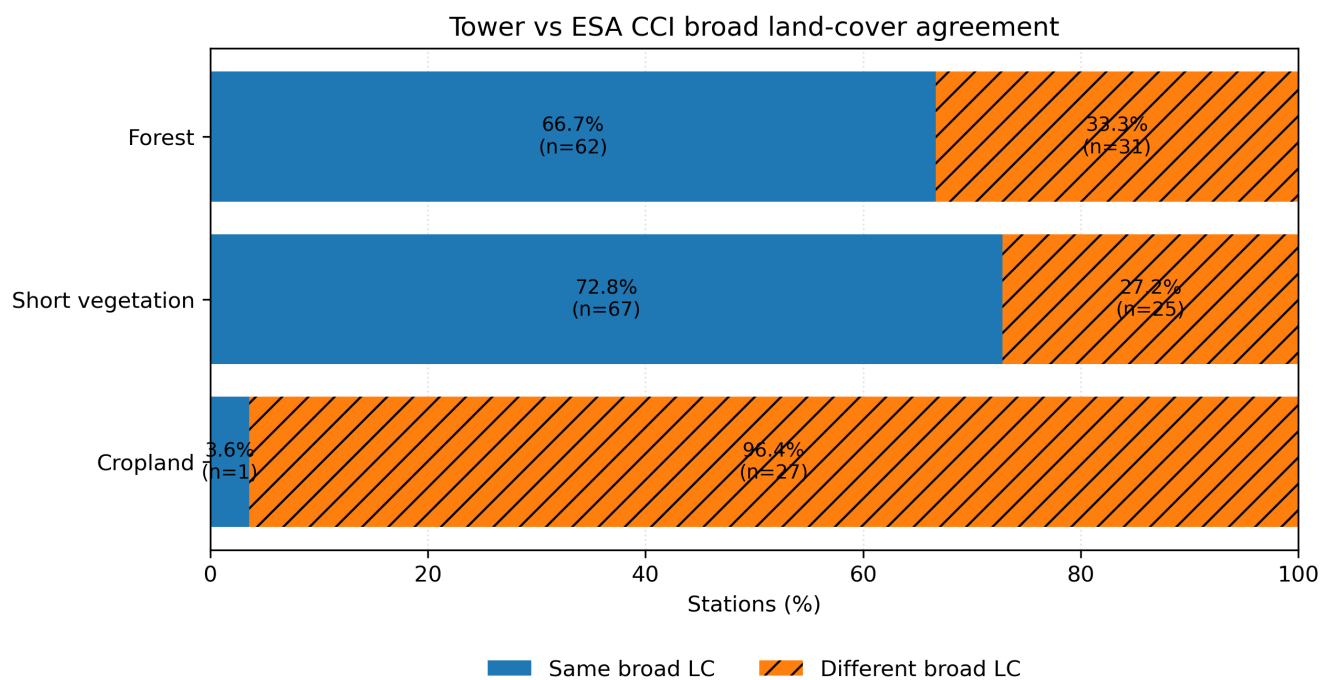


Figure A2. Tower–satellite broad LC agreement across vegetation groups. Percentage of flux-tower sites for which the tower LC class agrees with the corresponding ESA CCI broad LC class, shown separately for forest, short vegetation, and cropland sites.

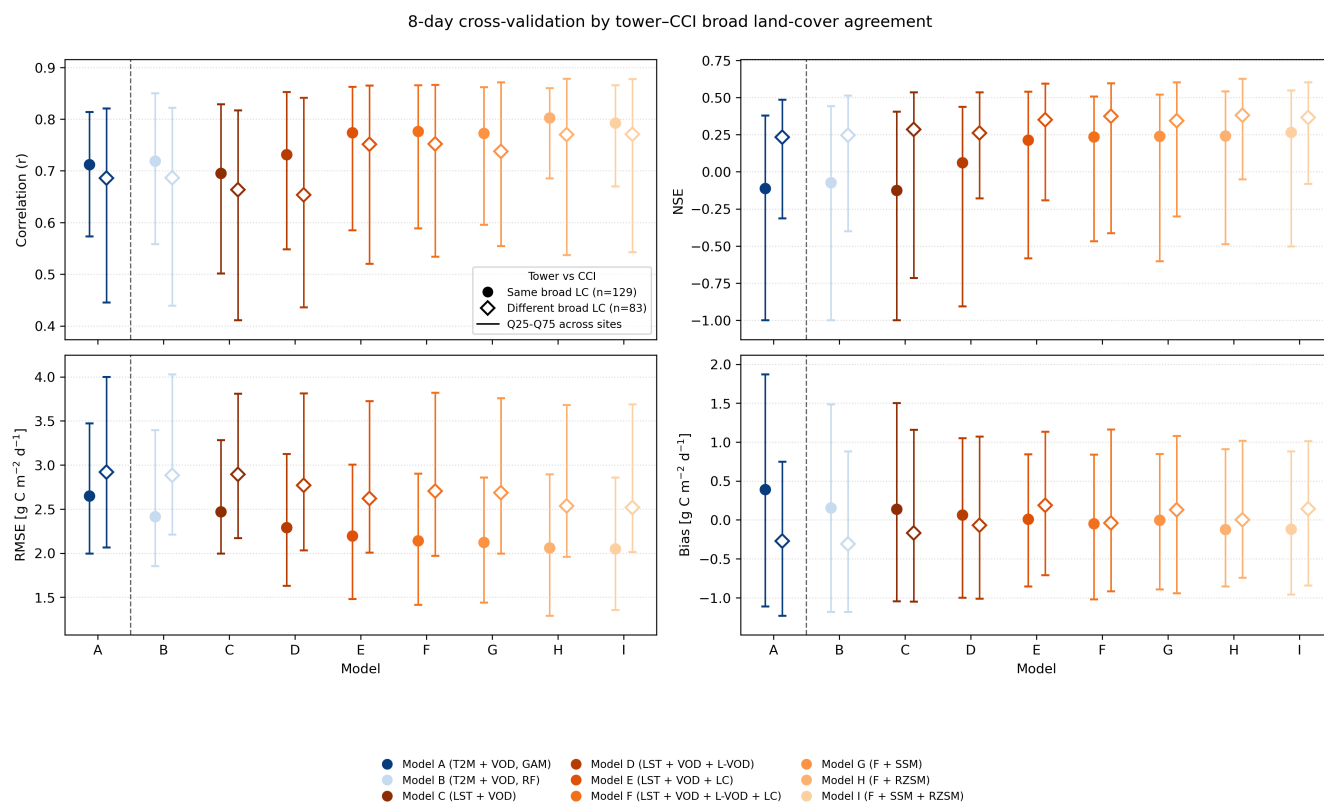


Figure A3. Cross-validation performance stratified by tower-satellite LC agreement. Sites are classified according to whether the dominant LC class at the flux-tower location agrees with that of the corresponding satellite grid cell.

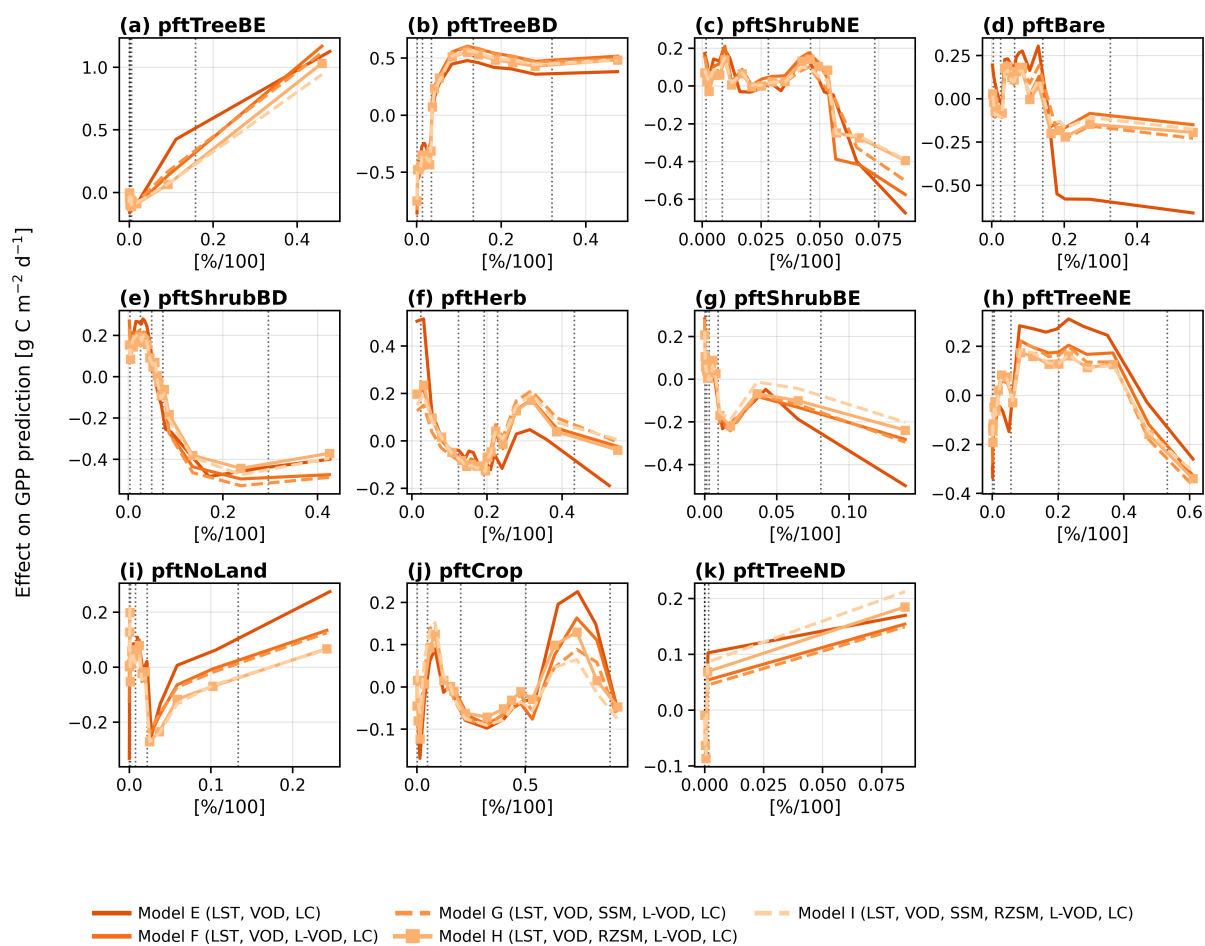


Figure A4. Accumulated Local Effects (ALE) of plant functional type (PFT) predictors across all models. Curves show centred 1D ALE (20 quantile bins) of GPP vs each PFT fraction; dotted vertical lines mark the 0.05/0.25/0.50/0.75/0.95 quantiles of the reference model's PFT distribution. Positive/negative ALE values indicate model effects above/below the centred mean effect over the data used for the ALE calculation.

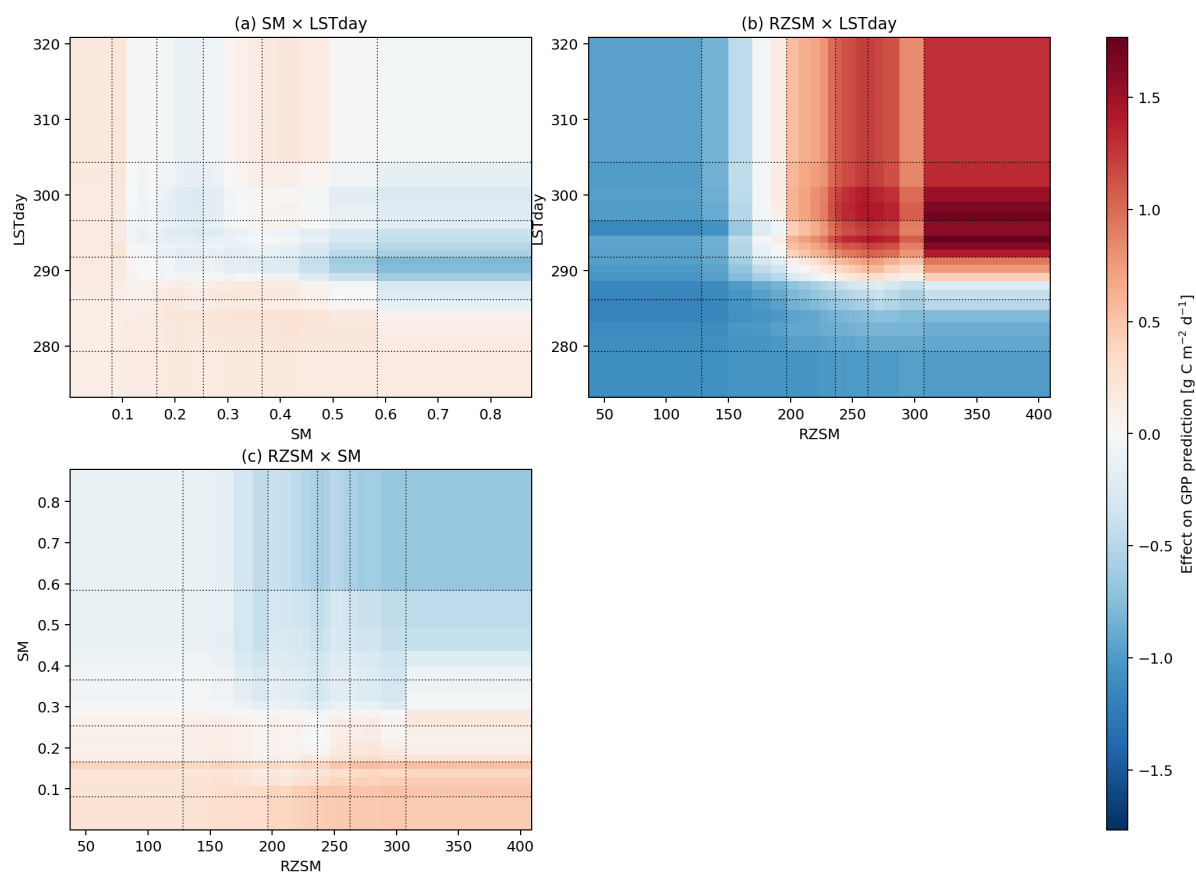


Figure A5. 2D Accumulated Local Effects (ALE) for selected predictor pairs. Positive/negative values indicate positive/negative second-order interaction effects relative to the additive main effects of the two predictors.



Table A4. Mean annual GPP over the common spatial domain (115.1 million km²) and the common period 2003–2015. For Models A, B, H, MODIS, FLUXCOM-RS, and FLUXCOM X-Base, values are reported as the temporal mean $\pm$ interannual standard deviation. For TRENDY-v7, the ensemble mean is followed by the minimum–maximum range across the mean GPP estimates of the 16 ensemble members.

	Configuration / product	GPP (Pg C yr ⁻¹)
Models	Model A T2M, VOD (GAM)	206.8 $\pm$ 2.2
	Model B T2M, VOD (RF)	189.4 $\pm$ 1.1
	Model H LST, VOD, RZSM, L-VOD, LC	175.0 $\pm$ 1.4
References	MODIS	111.1 $\pm$ 2.7
	FLUXCOM-RS	108.4 $\pm$ 1.3
	FLUXCOM X-Base	121.5 $\pm$ 2.1
	TRENDY-v7	116.4 [72.1–148.0]

Table A5. Global and regional relative inter-fold GPP uncertainty, expressed as the coefficient of variation (CV = SD/mean, in %), for Models B, and G.

Model	Global mean CV [%]	Sahara/Sahel [%]	Australia (arid) [%]	Boreal (>50N) [%]
B	14.2	21.1	24.8	10.4
H	7.9	14.7	20.3	4.8

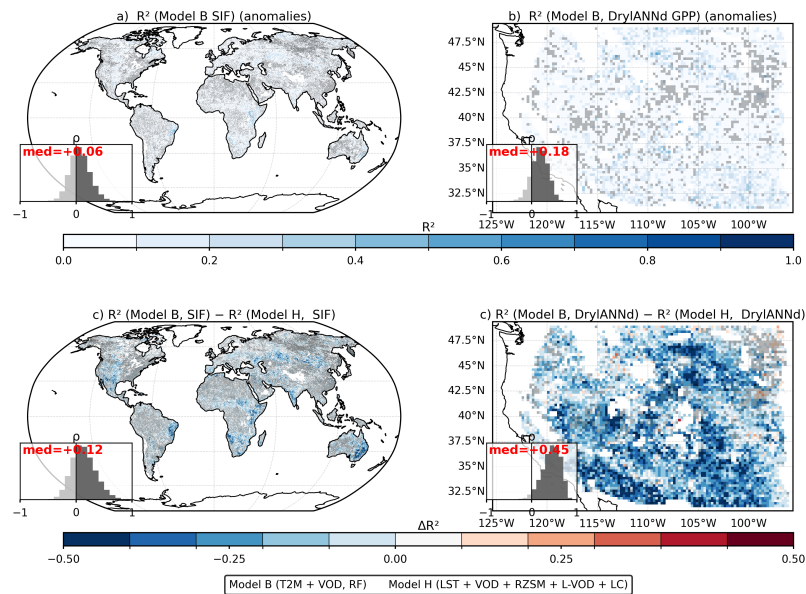


Figure A6. Spatial agreement over anomalies of Models B and H with SIF (left) and DrylANNd GPP (right). The upper panels show R^2 for Model B, while the lower panels show the difference between Models B and H. Blue indicates higher R^2 for Model H and red higher R^2 for Model B. Grey marks cells with negative Spearman ρ or fewer than 24 valid collocated observations. Insets show the distribution of Spearman ρ and its median.

Table A6. Agreement between predicted GPP and SIF, for the raw (full) signal and for anomalies. R^2 and ρ (Spearman) are unweighted medians over all grid cells passing the minimum-sample threshold. “% $\rho < 0$ ” is the unweighted fraction of those cells with a negative rank correlation.

Model	R^2 full	R^2 anom	ρ full	ρ anom	% $\rho < 0$ full	% $\rho < 0$ anom
Model B	0.178	0.004	0.422	0.064	17.2	35.6
Model H	0.311	0.015	0.557	0.124	13.3	27.0

Table A7. Agreement between predicted GPP and DrylANNd, for the raw (full) signal and for anomalies. R^2 and ρ (Spearman) are unweighted medians over all grid cells passing the minimum-sample threshold. “% $\rho < 0$ ” is the unweighted fraction of those cells with a negative rank correlation.

Model	R^2 full	R^2 anom	ρ full	ρ anom	% $\rho < 0$ full	% $\rho < 0$ anom
Model B	0.609	0.034	0.780	0.183	3.3	16.6
Model H	0.707	0.205	0.841	0.452	1.7	2.8

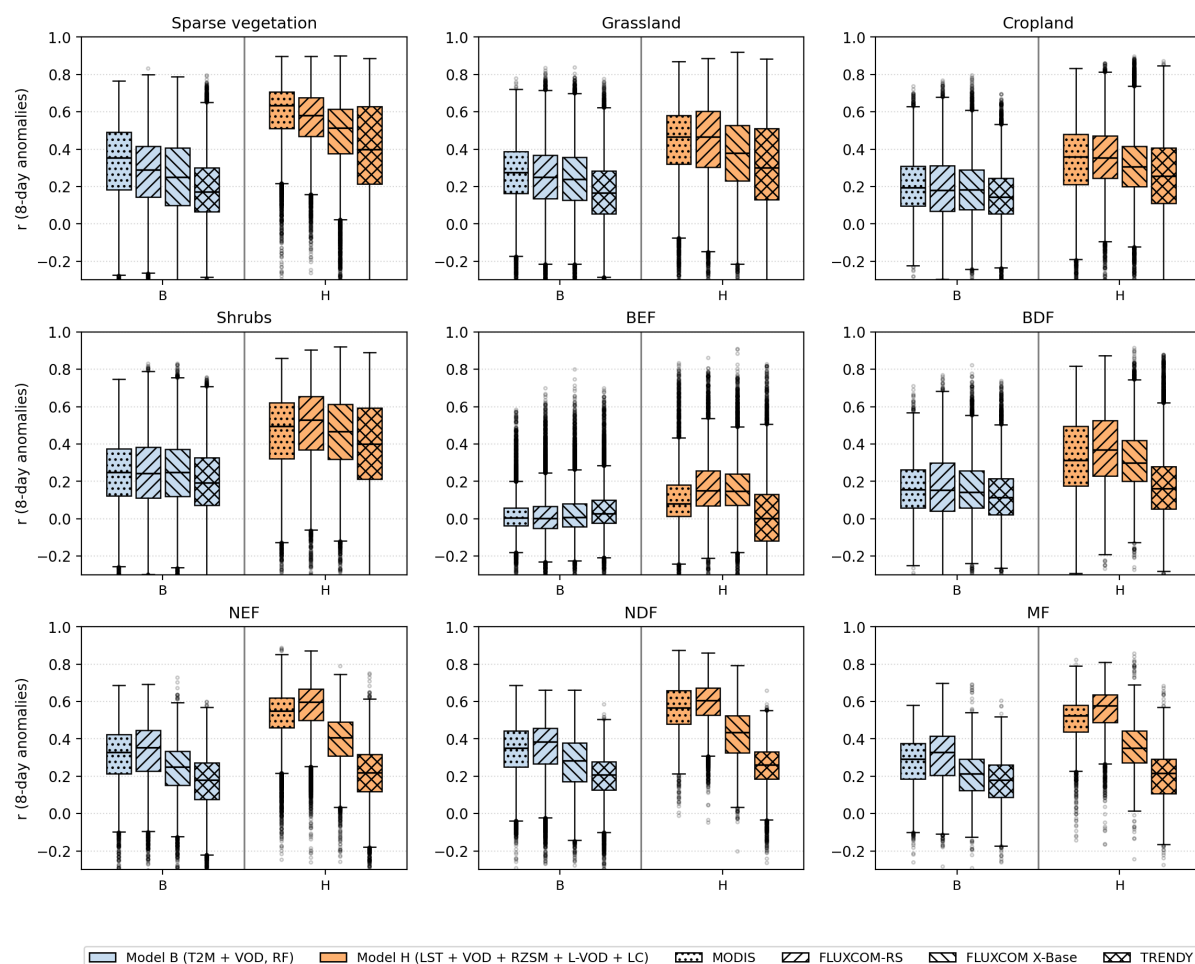


Figure A7. Pixel-wise correlation (r) between predicted GPP and four independent references (FLUXCOM RS, FLUXCOM X-Base, MODIS, TRENDY) computed on 8-day anomalies, shown as boxplots by land-cover class. Anomalies are derived by subtracting an 8-day-of-year climatology to remove the seasonal cycle, isolating sub-seasonal and interannual co-variability. Correlations are masked where fewer than 24 paired 8-day samples are available.



Table A8. Median pixel-wise Pearson correlation coefficient (r) between GPP from Models B and H and four external GPP products, stratified by land-cover class. Panel (a) shows correlations for the full 8-day signal and Panel (b) for 8-day anomalies.

LC class	MODIS		FLUXCOM-RS		FLUXCOM X-Base		TRENDY	
	B	H	B	H	B	H	B	H
<i>(a) Full 8-day signal</i>								
Sparse vegetation	0.570	0.669	0.552	0.716	0.482	0.630	0.495	0.677
Grassland	0.705	0.784	0.684	0.788	0.693	0.761	0.651	0.751
Cropland	0.673	0.788	0.657	0.796	0.647	0.776	0.636	0.813
Shrubs	0.565	0.754	0.581	0.805	0.600	0.786	0.549	0.753
BEF	0.024	0.110	-0.012	0.264	0.001	0.241	0.083	0.050
BDF	0.708	0.835	0.716	0.846	0.721	0.835	0.726	0.840
NEF	0.771	0.841	0.815	0.875	0.794	0.835	0.736	0.755
NDF	0.783	0.855	0.814	0.874	0.782	0.828	0.776	0.804
MF	0.814	0.859	0.861	0.878	0.839	0.845	0.807	0.798
<i>(b) 8-day anomalies</i>								
Sparse vegetation	0.355	0.634	0.289	0.582	0.249	0.514	0.172	0.398
Grassland	0.275	0.467	0.252	0.465	0.240	0.378	0.167	0.301
Cropland	0.196	0.359	0.181	0.355	0.184	0.306	0.143	0.257
Shrubs	0.248	0.497	0.244	0.528	0.249	0.466	0.194	0.399
BEF	0.006	0.081	0.001	0.150	0.009	0.147	0.028	0.003
BDF	0.156	0.313	0.152	0.369	0.144	0.300	0.116	0.162
NEF	0.329	0.550	0.355	0.599	0.250	0.406	0.181	0.219
NDF	0.352	0.566	0.384	0.606	0.283	0.435	0.208	0.260
MF	0.291	0.523	0.328	0.578	0.213	0.350	0.180	0.216



Author contributions. Author contribution text (mandatory).

Ruxandra-Maria Zotta: Writing – original draft, methodology, formal analysis, data curation, conceptualization. Moritz Clemens
Müller: review and editing. Raul Lezameta: review and editing. Sophia Walther: review, conceptualisation, data curation.
Matthias Forkel: review, conceptualisation. Wouter Dorigo: writing – review and editing, supervision, methodology, funding
655 acquisition, conceptualization.

Competing interests. The authors declare no conflict of interest.

Disclaimer. Optional disclaimer text.

Acknowledgements. The authors acknowledge the TU Wien University Library for financial support through its Open Access Funding Programme.

660 During manuscript preparation, ChatGPT (OpenAI) and Grammarly were used for language editing and text correction. Claude (Anthropic) was used to assist with code development and debugging; all outputs were reviewed and verified by the authors.



References

- Ahlström, A., Raupach, M. R., Schurgers, G., Smith, B., Arneth, A., Jung, M., Reichstein, M., Canadell, J. G., Friedlingstein, P., Jain, A. K., et al.: The dominant role of semi-arid ecosystems in the trend and variability of the land CO₂ sink, *Science*, 348, 895–899, 2015.
- 665 Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., Murray-Tortarolo, G., Papale, D., Parazoo, N. C., Peylin, P., et al.: Spatiotemporal patterns of terrestrial gross primary production: A review, *Reviews of Geophysics*, 53, 785–818, 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, 2020.
- Atkin, O. K. and Tjoelker, M. G.: Thermal acclimation and the dynamic response of plant respiration to temperature, *Trends in plant science*, 8, 343–351, 2003.
- 670 Baldocchi, D.: Measuring fluxes of trace gases and energy between ecosystems and the atmosphere—the state and future of the eddy covariance method, *Global change biology*, 20, 3600–3609, 2014.
- Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M. A., Baldocchi, D., Bonan, G. B., et al.: Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate, *Science*, 329, 834–838, 2010.
- 675 Berg, W., Kroodsma, R., Kummerow, C. D., and McKague, D. S.: Fundamental climate data records of microwave brightness temperatures, *Remote Sensing*, 10, 1306, 2018.
- Boston College Institute for Scientific Research: General Characteristics of DMSP — Equator Crossing Times, <https://dmsp.bc.edu/html2/ssiesdmspgeneral.html>, 2025.
- Breiman, L.: Random forests, *Machine learning*, 45, 5–32, 2001.
- 680 CCI, E. L. C.: Product user guide version 2.0, UCL-Geomatics: London, UK, 685, 2017.
- Chang, Q., Xiao, X., Doughty, R., Wu, X., Jiao, W., and Qin, Y.: Assessing variability of optimum air temperature for photosynthesis across site-years, sites and biomes and their effects on photosynthesis estimation, *Agricultural and Forest Meteorology*, 298, 108 277, 2021.
- Chu, H., Luo, X., Ouyang, Z., Chan, W. S., Dengel, S., Biraud, S. C., Torn, M. S., Metzger, S., Kumar, J., Arain, M. A., et al.: Representativeness of Eddy-Covariance flux footprints for areas surrounding AmeriFlux sites, *Agricultural and Forest Meteorology*, 301, 108 350, 2021.
- 685 Collalti, A. and Prentice, I.: Is NPP proportional to GPP? Waring’s hypothesis 20 years on, *Tree physiology*, 39, 1473–1483, 2019.
- Cuntz, M., Smith, B., Canadell, J. G., Knauer, J., and Haverd, V.: Large and increasing biospheric productivity of northern ecosystems, *Geophysical Research Letters*, 52, e2025GL115 983, 2025.
- Dannenber, 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, 2023.
- 690 de Nijs, A. H., Parinussa, R. M., de Jeu, R. A., Schellekens, J., and Holmes, T. R.: A methodology to determine radio-frequency interference in AMSR2 observations, *IEEE Transactions on Geoscience and Remote Sensing*, 53, 5148–5159, 2015.
- Defourny, P. and ESA Land Cover CCI Project Team: ESA Land Cover Climate Change Initiative (Land_Cover_cci): Global Land Cover Maps, Version 2.0.7, Dataset, version 2.0.7, 2017.
- 695 Delwiche, K. B., Knox, S. H., Malhotra, A., Fluet-Chouinard, E., McNicol, G., Feron, S., Ouyang, Z., Papale, D., Trotta, C., Canfora, E., et al.: FLUXNET-CH4: A global, multi-ecosystem dataset and analysis of methane seasonality from freshwater wetlands, *Earth System Science Data Discussions*, 2021, 1–111, 2021.



- Donlon, C., Galeazzi, C., Mithassel, R., Sallusti, M., Triggianese, M., Fiorelli, B., De Paris, G., Kornienko, A., and Khlystova, I.: The
 700 Copernicus imaging microwave radiometer (CIMR): Mission overview and status, in: IGARSS 2023-2023 IEEE International Geoscience
 and Remote Sensing Symposium, pp. 989–992, IEEE, 2023.
- Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., et al.: ESA CCI Soil
 Moisture for improved Earth system understanding: State-of-the art and future directions, *Remote Sensing of Environment*, 203, 185–215,
 2017.
- 705 Dorigo, W., Preimesberger, W., Lems, J., Frederikse, T., and Dostalova, A.: Soil moisture gridded data from 1978 to present,
<https://doi.org/10.24381/cds.d7782f18>, accessed on YYYY-MM-DD, 2018.
- European Commission, Joint Research Centre (JRC): GDO Standardized Precipitation-Evapotranspiration Index ERA5, 1-
 month accumulation period (SPEI-1), version 1.0.0, <http://data.europa.eu/89h/334bab8f-3467-4d89-8b7a-536a06227a7e>, pID:
<http://data.europa.eu/89h/334bab8f-3467-4d89-8b7a-536a06227a7e>, 2025a.
- 710 European Commission, Joint Research Centre (JRC): GDO Standardized Precipitation-Evapotranspiration Index ERA5, 3-
 month accumulation period (SPEI-3), version 1.0.0, <http://data.europa.eu/89h/e3403941-2a44-41a0-9698-c1f7fe60e3c0>, pID:
<http://data.europa.eu/89h/e3403941-2a44-41a0-9698-c1f7fe60e3c0>, 2025b.
- European Commission, Joint Research Centre (JRC): GDO Standardized Precipitation-Evapotranspiration Index ERA5, 6-
 month accumulation period (SPEI-6), version 1.0.0, <http://data.europa.eu/89h/4ce50544-d434-4897-92ab-a9b720e1a052>, pID:
 715 <http://data.europa.eu/89h/4ce50544-d434-4897-92ab-a9b720e1a052>, 2025c.
- Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B., and Otero-Casal, C.: Hydrologic regulation of plant rooting depth, *Proceedings
 of the National Academy of Sciences*, 114, 10 572–10 577, 2017.
- Fang, J., Fang, J., Chen, B., Zhang, H., Dilawar, A., Guo, M., and Liu, S.: Assessing Spatial Representativeness of Global Flux Tower Eddy-
 Covariance Measurements Using Data from FLUXNET2015, *Scientific Data*, 11, 569, <https://doi.org/10.1038/s41597-024-03291-3>, 2024.
- 720 Frappart, F., Wigneron, J.-P., Li, X., Liu, X., Al-Yaari, A., Fan, L., Wang, M., Moisy, C., Le Masson, E., Aoulad Lafkih, Z., et al.: Global
 monitoring of the vegetation dynamics from the vegetation optical depth (VOD): A review, *Remote Sensing*, 12, 2915, 2020.
- Friedlingstein, P., O'sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Landschützer, P., Le Quéré, C., Li, H., Luijkx, I. T., Olsen, A.,
 et al.: Global carbon budget 2024, *Earth System Science Data Discussions*, 2024, 1–133, 2024.
- Good, E. J., Ghent, D. J., Bulgin, C. E., and Remedios, J. J.: A spatiotemporal analysis of the relationship between near-surface air temperature
 725 and satellite land surface temperatures using 17 years of data from the ATSR series, *Journal of Geophysical Research: Atmospheres*, 122,
 9185–9210, 2017.
- Green, J. K., Seneviratne, S. I., Berg, A. M., Findell, K. L., Hagemann, S., Lawrence, D. M., and Gentine, P.: Large influence of soil moisture
 on long-term terrestrial carbon uptake, *Nature*, 565, 476–479, 2019.
- Grossiord, C., Buckley, T. N., Cernusak, L. A., Novick, K. A., Poulter, B., Siegwolf, R. T., Sperry, J. S., and McDowell, N. G.: Plant responses
 730 to rising vapor pressure deficit, *New phytologist*, 226, 1550–1566, 2020.
- Guan, K., Pan, M., Li, H., Wolf, A., Wu, J., Medvigy, D., Caylor, K. K., Sheffield, J., Wood, E. F., Malhi, Y., et al.: Photosynthetic seasonality
 of global tropical forests constrained by hydroclimate, *Nature Geoscience*, 8, 284–289, 2015.
- Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, J. A., Frankenberg, C., Huete, A. R., Zarco-Tejada, P., Lee, J.-E., et al.: Global
 and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence, *Proceedings of the National Academy of Sciences*,
 735 111, E1327–E1333, 2014.



- Guanter, L., Bacour, C., Schneider, A., Aben, I., van Kempen, T. A., Maignan, F., Retscher, C., Köhler, P., Frankenberg, C., Joiner, J., et al.: The TROPIS global sun-induced fluorescence dataset from the Sentinel-5P TROPOMI mission, *Earth System Science Data*, 13, 5423–5440, 2021.
- Güntner, A., Sharifi, E., Haas, J., Boergens, E., Cao, F., Dahle, C., Darbeheshti, N., Dobslaw, H., Dussaillant, I., Dorigo, W., et al.: The
740 Global Gravity-based Groundwater Product (G3P), *Earth System Science Data Discussions*, 2026, 1–48, 2026.
- Holmes, T., De Jeu, R., Owe, M., and Dolman, A.: Land surface temperature from Ka band (37 GHz) passive microwave observations, *Journal of Geophysical Research: Atmospheres*, 114, 2009.
- Humphrey, V., Zscheischler, J., Ciais, P., Gudmundsson, L., Sitch, S., and Seneviratne, S. I.: Sensitivity of atmospheric CO₂ growth rate to observed changes in terrestrial water storage, *Nature*, 560, 628–631, 2018.
- 745 Humphrey, V., Berg, A., Ciais, P., Gentine, P., Jung, M., Reichstein, M., Seneviratne, S. I., and Frankenberg, C.: Soil moisture–atmosphere feedback dominates land carbon uptake variability, *Nature*, 592, 65–69, 2021.
- Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., et al.: Compensatory water effects link yearly global land CO₂ sink changes to temperature, *Nature*, 541, 516–520, 2017.
- Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N.,
750 et al.: Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach, *Biogeosciences*, 17, 1343–1365, 2020.
- Kannenber, S. A., Anderegg, W. R., Barnes, M. L., Dannenberg, M. P., and Knapp, A. K.: Dominant role of soil moisture in mediating carbon and water fluxes in dryland ecosystems, *Nature Geoscience*, 17, 38–43, 2024.
- Kawanishi, T., Sezai, T., Ito, Y., Imaoka, K., Takeshima, T., Ishido, Y., Shibata, A., Miura, M., Inahata, H., and Spencer, R. W.: The Advanced
755 Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), NASA’s contribution to the EOS for global energy and water cycle studies, *IEEE Transactions on Geoscience and Remote Sensing*, 41, 184–194, 2003.
- Knowles, K., Savoie, M., Armstrong, R., and Brodzik, M.: AMSR-E/Aqua Daily EASE-Grid Brightness Temperatures, Version 1, Boulder, Colorado USA, NASA National Snow and Ice Data Center Distributed Active Archive Center [data set], 2006.
- Koirala, S., Jung, M., Reichstein, M., de Graaf, I. E., Camps-Valls, G., Ichii, K., Papale, D., Ráduly, B., Schwalm, C. R., Tramontana, G.,
760 et al.: Global distribution of groundwater-vegetation spatial covariation, *Geophysical Research Letters*, 44, 4134–4142, 2017.
- Konings, A. G. and Gentine, P.: Global variations in ecosystem-scale isohydricity, *Global change biology*, 23, 891–905, 2017.
- Konings, A. G., Rao, K., and Steele-Dunne, S. C.: Macro to micro: microwave remote sensing of plant water content for physiology and ecology, *New Phytologist*, 223, 1166–1172, 2019.
- Konings, A. G., Holtzman, N. M., Rao, K., Xu, L., and Saatchi, S. S.: Interannual variations of vegetation optical depth are due to both water
765 stress and biomass changes, *Geophysical Research Letters*, 48, e2021GL095 267, 2021.
- Kuhn-Régnier, A., Voulgarakis, A., Nowack, P., Forkel, M., Prentice, I. C., and Harrison, S. P.: The importance of antecedent vegetation and drought conditions as global drivers of burnt area, *Biogeosciences*, 18, 3861–3879, 2021.
- Lasslop, G., Reichstein, M., Papale, D., Richardson, A. D., Arneth, A., Barr, A., Stoy, P., and Wohlfahrt, G.: Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation, *Global change
770 biology*, 16, 187–208, 2010.
- Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., Pickers, P. A., Korsbakken, J. I., Peters, G. P., Canadell, J. G., et al.: Global carbon budget 2018, *Earth System Science Data*, 10, 2141–2194, 2018.



- Li, W., Migliavacca, M., Forkel, M., Walther, S., Reichstein, M., and Orth, R.: Revisiting global vegetation controls using multi-layer soil moisture, *Geophysical Research Letters*, 48, e2021GL092856, 2021a.
- 775 Li, X. and Xiao, J.: TROPOMI observations allow for robust exploration of the relationship between solar-induced chlorophyll fluorescence and terrestrial gross primary production, *Remote Sensing of Environment*, 268, 112748, 2022.
- Li, X., Xiao, J., Fisher, J. B., and Baldocchi, D. D.: ECOSTRESS estimates gross primary production with fine spatial resolution for different times of day from the International Space Station, *Remote Sensing of Environment*, 258, 112360, 2021b.
- Lin, Y.-S., Medlyn, B. E., and Ellsworth, D. S.: Temperature responses of leaf net photosynthesis: the role of component processes, *Tree*
- 780 *physiology*, 32, 219–231, 2012.
- Liu, L., Gudmundsson, L., Hauser, M., Qin, D., Li, S., and Seneviratne, S. I.: Soil moisture dominates dryness stress on ecosystem production globally, *Nature communications*, 11, 4892, 2020.
- Liu, Y., Peñuelas, J., Cescatti, A., Zhang, Y., and Zhang, Z.: Atmospheric dryness dominates afternoon depression of global terrestrial photosynthesis, *Geophysical Research Letters*, 51, e2024GL110954, 2024.
- 785 Liu, Y. Y., De Jeu, R. A., McCabe, M. F., Evans, J. P., and Van Dijk, A. I.: Global long-term passive microwave satellite-based retrievals of vegetation optical depth, *Geophysical Research Letters*, 38, 2011.
- 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.: GLEAM v3: Satellite-based land evaporation and root-zone soil moisture, *Geoscientific Model Development*, 10, 1903–1925, 2017.
- Meesters, A. G., De Jeu, R. A., and Owe, M.: Analytical derivation of the vegetation optical depth from the microwave polarization difference
- 790 *index*, *IEEE Geoscience and Remote Sensing Letters*, 2, 121–123, 2005.
- Meier, W., Jesefino, C., and Thorsten, M.: NRT AMSR2 Unified L3 Daily 25 km Brightness Temperature & Sea Ice Concentration Polar Grids V1, 2018.
- Mialon, A., Rodríguez-Fernández, N. J., Santoro, M., Saatchi, S., Mermoz, S., Bousquet, E., and Kerr, Y. H.: Evaluation of the sensitivity of SMOS L-VOD to forest above-ground biomass at global scale, *Remote Sensing*, 12, 1450, 2020.
- 795 Miller, D. L., Wolf, S., Fisher, J. B., Zaitchik, B. F., Xiao, J., and Keenan, T. F.: Increased photosynthesis during spring drought in energy-limited ecosystems, *Nature Communications*, 14, 7828, 2023.
- Moesinger, L., Dorigo, W., De Jeu, R., Van Der Schalie, R., Scanlon, T., Teubner, I., and Forkel, M.: The global long-term microwave vegetation optical depth climate archive (VODCA), *Earth System Science Data*, 12, 177–196, 2020.
- Momen, M., Wood, J. D., Novick, K. A., Pangle, R., Pockman, W. T., McDowell, N. G., and Konings, A. G.: Interacting effects of leaf water
- 800 *potential and biomass on vegetation optical depth*, *Journal of Geophysical Research: Biogeosciences*, 122, 3031–3046, 2017.
- Monteith, J. L.: Solar radiation and productivity in tropical ecosystems, *Journal of applied ecology*, 9, 747–766, 1972.
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., et al.: ERA5-Land: A state-of-the-art global reanalysis dataset for land applications, *Earth system science data*, 13, 4349–4383, 2021.
- 805 Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I—A discussion of principles, *Journal of hydrology*, 10, 282–290, 1970.
- Nelson & Walther, Gans, F., Kraft, B., Weber, U., Novick, K., Buchmann, N., Migliavacca, M., Wohlfahrt, G., Šigut, L., et al.: X-BASE: the first terrestrial carbon and water flux products from an extended data-driven scaling framework, *FLUXCOM-X, Biogeosciences*, 21, 5079–5115, 2024.



- 810 Nemani, R. R., Keeling, C. D., Hashimoto, H., Jolly, W. M., Piper, S. C., Tucker, C. J., Myneni, R. B., and Running, S. W.: Climate-driven increases in global terrestrial net primary production from 1982 to 1999, *science*, 300, 1560–1563, 2003.
- NSIDC: AMSR-E Instrument Description, <https://nsidc.org/sites/default/files/amser-e20instrument20description.pdf>, specifies AMSR-E overpass times near 01:30 and 13:30 local time, 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., et al.: Climate-driven variability and trends in plant productivity over recent decades based on three global products, *Global Biogeochemical Cycles*, 34, e2020GB006613, 2020.
- 815 Owe, M., de Jeu, R., and Walker, J.: A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index, *IEEE Transactions on Geoscience and Remote Sensing*, 39, 1643–1654, 2002.
- Parinussa, R., Holmes, T., Yilmaz, M., and Crow, W.: The impact of land surface temperature on soil moisture anomaly detection from passive microwave observations, *Hydrology and Earth System Sciences*, 15, 3135–3151, 2011.
- 820 Pasik, A., Gruber, A., Preimesberger, W., De Santis, D., and Dorigo, W.: Uncertainty estimation for a new exponential-filter-based long-term root-zone soil moisture dataset from Copernicus Climate Change Service (C3S) surface observations, *Geoscientific Model Development*, 16, 4957–4976, 2023.
- Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., et al.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, *Scientific data*, 7, 225, 2020.
- 825 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: Machine learning in Python, the *Journal of machine Learning research*, 12, 2825–2830, 2011.
- Poulter, B., Frank, D., Ciais, P., Myneni, R. B., Andela, N., Bi, J., Broquet, G., Canadell, J. G., Chevallier, F., Liu, Y. Y., et al.: Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle, *Nature*, 509, 600–603, 2014.
- 830 Poulter, B., MacBean, N., Hartley, A., Khlystova, I., Arino, O., Betts, R., Bontemps, S., Boettcher, M., Brockmann, C., Defourny, P., et al.: Plant functional type classification for earth system models: Results from the European Space Agency’s Land Cover Climate Change Initiative, *Geoscientific Model Development*, 8, 2315–2328, 2015.
- Preimesberger, W., Stradiotti, P., and Dorigo, W. A.: ESA CCI SM GAPFILLED Long-term Climate Data Record of Surface Soil Moisture from merged multi-satellite observations, <https://doi.org/10.48436/s5j4q-rpd32>, dataset; published 5 July 2024; modified 27 January 2025.
- 835 License: CC-BY-4.0, 2024.
- Preimesberger, W., Stradiotti, P., and Dorigo, W.: ESA CCI Soil Moisture GAPFILLED: an independent global gap-free satellite climate data record with uncertainty estimates, *Earth System Science Data*, 17, 4305–4329, 2025.
- Qiu, R., Han, G., Li, X., Xiao, J., Liu, J., Wang, S., Li, S., and Gong, W.: Contrasting responses of relationship between solar-induced fluorescence and gross primary production to drought across aridity gradients, *Remote Sensing of Environment*, 302, 113984, 2024.
- 840 Quegan, S., Le Toan, T., Chave, J., et al.: The European Space Agency BIOMASS mission: Measuring forest above-ground biomass from space, *Remote Sensing of Environment*, 227, 44–60, <https://doi.org/10.1016/j.rse.2019.03.032>, 2019.
- Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., et al.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, *Global change biology*, 11, 1424–1439, 2005.
- 845 Remote Sensing Systems: AMSR Instrument Series — Equator Crossing Times and Mission Specs, <https://www.remss.com/missions/amser/>, 2025.



- Restrepo-Coupe, N., da Rocha, H. R., Hutyrá, L. R., da Araujo, A. C., Borma, L. S., Christoffersen, B., Cabral, O. M., de Camargo, P. B., Cardoso, F. L., da Costa, A. C. L., et al.: What drives the seasonality of photosynthesis across the Amazon basin? A cross-site analysis of eddy flux tower measurements from the Brasil flux network, *Agricultural and Forest Meteorology*, 182, 128–144, 2013.
- 850 Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillerá-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., et al.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, *Ecography*, 40, 913–929, 2017.
- Rodríguez-Fernández, N. J., Mialon, A., Mermoz, S., Bouvet, A., Richaume, P., Al Bitar, A., Al-Yaari, A., Brandt, M., Kaminski, T., Le Toan, T., et al.: An evaluation of SMOS L-band vegetation optical depth (L-VOD) data sets: high sensitivity of L-VOD to above-ground biomass in Africa, *Biogeosciences*, 15, 4627–4645, 2018.
- 855 Roxburgh, S., Berry, S. L., Buckley, T., Barnes, B., and Roderick, M.: What is NPP? Inconsistent accounting of respiratory fluxes in the definition of net primary production, *Functional Ecology*, 19, 378–382, 2005.
- Running, S., Mu, Q., and Zhao, M.: MOD17A3H MODIS/terra net primary production yearly L4 global 500m SIN grid V006, NASA EOSDIS Land Processes Distributed Active Archive Center (DAAC) data set, pp. MOD17A3H–006, 2015.
- 860 Running, S. W. and Zhao, M.: User's guide daily GPP and annual NPP (MOD17A2H/A3H) and year-end gap-filled (MOD17A2HGF/A3HGF) products NASA earth observing system MODIS land algorithm (for collection 6), Process. DAAC, 490, 1–37, 2019.
- Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., and Hashimoto, H.: A continuous satellite-derived measure of global terrestrial primary production, *Bioscience*, 54, 547–560, 2004.
- 865 Schickling, A., Matveeva, M., Damm, A., Schween, J. H., Wahner, A., Graf, A., Crewell, S., and Rascher, U.: Combining sun-induced chlorophyll fluorescence and photochemical reflectance index improves diurnal modeling of gross primary productivity, *Remote Sensing*, 8, 574, 2016.
- Schmidt, L., Forkel, M., Zotta, R.-M., Scherrer, S., Dorigo, W. A., Kuhn-Régner, A., Van Der Schalie, R., and Yebra, M.: Assessing the sensitivity of multi-frequency passive microwave vegetation optical depth to vegetation properties, *Biogeosciences Discussions*, 2022, 1–24, 2022.
- 870 Sitch, S., Friedlingstein, P., Gruber, N., Jones, S. D., Murray-Tortarolo, G., Ahlström, A., Doney, S. C., Graven, H., Heinze, C., Huntingford, C., et al.: Recent trends and drivers of regional sources and sinks of carbon dioxide, *Biogeosciences*, 12, 653–679, 2015.
- Still, C. J., Rastogi, B., Page, G. F., Griffith, D. M., Sibley, A., Schulze, M., Hawkins, L., Pau, S., Detto, M., and Helliker, B. R.: Imaging canopy temperature: shedding (thermal) light on ecosystem processes, *New Phytologist*, 230, 1746–1753, 2021.
- 875 Sulkava, M., Luyssaert, S., Zaehle, S., and Papale, D.: Assessing and improving the representativeness of monitoring networks: The European flux tower network example, *Journal of Geophysical Research: Biogeosciences*, 116, 2011.
- Teubner, I. E., Forkel, M., Jung, M., Liu, Y. Y., Miralles, D. G., Parinussa, R., Van der Schalie, R., Vreugdenhil, M., Schwalm, C. R., Tramontana, G., et al.: Assessing the relationship between microwave vegetation optical depth and gross primary production, *International journal of applied earth observation and geoinformation*, 65, 79–91, 2018.
- 880 Teubner, I. E., Forkel, M., Camps-Valls, G., Jung, M., Miralles, D. G., Tramontana, G., Van der Schalie, R., Vreugdenhil, M., Möisinger, L., and Dorigo, W. A.: A carbon sink-driven approach to estimate gross primary production from microwave satellite observations, *Remote Sensing of Environment*, 229, 100–113, 2019.
- Teubner, I. E., Forkel, M., Wild, B., Möisinger, L., and Dorigo, W. A.: Impact of temperature and water availability on microwave-derived gross primary production, *Biogeosciences Discussions*, 2020, 1–28, 2020.



- 885 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, 2015.
- Tramontana, G., Jung, M., Schwalm, C. R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M. A., Cescatti, A., Kiely, G., et al.: Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, *Biogeosciences*, 13, 4291–4313, 2016.
- 890 Turner, D. P., Ritts, W. D., Cohen, W. B., Gower, S. T., Running, S. W., Zhao, M., Costa, M. H., Kirschbaum, A. A., Ham, J. M., Saleska, S. R., et al.: Evaluation of MODIS NPP and GPP products across multiple biomes, *Remote sensing of environment*, 102, 282–292, 2006.
- van Der Schalie, R., de Jeu, R. A., Kerr, Y. H., Wigneron, J.-P., Rodríguez-Fernández, N. J., Al-Yaari, A., Parinussa, R. M., Mecklenburg, S., and Drusch, M.: The merging of radiative transfer based surface soil moisture data from SMOS and AMSR-E, *Remote Sensing of Environment*, 189, 180–193, 2017.
- 895 Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I.: A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index, *Journal of climate*, 23, 1696–1718, 2010.
- Vicente-Serrano, S. M., Gouveia, C., Camarero, J. J., Beguería, S., Trigo, R., López-Moreno, J. I., Azorín-Molina, C., Pasho, E., Lorenzo-Lacruz, J., Revuelto, J., et al.: Response of vegetation to drought time-scales across global land biomes, *Proceedings of the National Academy of Sciences*, 110, 52–57, 2013.
- 900 Walther, S., Duveiller, G., Jung, M., Guanter, L., Cescatti, A., and Camps-Valls, G.: Satellite observations of the contrasting response of trees and grasses to variations in water availability, *Geophysical Research Letters*, 46, 1429–1440, 2019.
- Wang, Y., Fu, Z., Hu, Z., and Niu, S.: Tracking global patterns of drought-induced productivity loss along severity gradient, *Journal of Geophysical Research: Biogeosciences*, 127, e2021JG006 753, 2022.
- Warm Winter 2020 Team, . I. E. T. C.: Warm winter 2020 ecosystem eddy covariance flux product for 73 stations in FLUXNET-archive format—Release 2022-1 (version 1.0), 2022.
- 905 Wild, B., Teubner, I., Moesinger, L., Zotta, R.-M., Forkel, M., van der Schalie, R., Sitch, S., and Dorigo, W.: VODCA2GPP—a new, global, long-term (1988–2020) gross primary production dataset from microwave remote sensing, *Earth System Science Data*, 14, 1063–1085, 2022.
- Yang, S., Weng, F., Yan, B., Sun, N., and Goldberg, M.: Special Sensor Microwave Imager (SSM/I) intersensor calibration using a simultaneous conical overpass technique, *Journal of Applied Meteorology and Climatology*, 50, 77–95, 2011.
- 910 Yang, T., Gong, H., and Zhang, Y.: Higher sensitivity of L-band vegetation optical depth to foliage than total aboveground biomass, *Geophysical Research Letters*, 52, e2025GL115 437, 2025.
- Yuan, Q., Wang, X., Che, T., and Li, J.: Global carbon flux dataset generated by fusing remote sensing and multiple flux networks observation, *Scientific Data*, 12, 1359, 2025.
- 915 Zhang, Y., Xiao, X., Jin, C., Dong, J., Zhou, S., Wagle, P., Joiner, J., Guanter, L., Zhang, Y., Zhang, G., et al.: Consistency between sun-induced chlorophyll fluorescence and gross primary production of vegetation in North America, *Remote Sensing of Environment*, 183, 154–169, 2016.
- Zhang, Y., Zhou, S., Gentile, P., and Xiao, X.: Can vegetation optical depth reflect changes in leaf water potential during soil moisture dry-down events?, *Remote Sensing of Environment*, 234, 111 451, 2019.
- 920 Zhang, Y., Feng, X., Fu, B., Chen, Y., and Wang, X.: Satellite-observed global terrestrial vegetation production in response to water availability, *Remote Sensing*, 13, 1289, 2021.



- Zhao, M., Heinsch, F. A., Nemani, R. R., and Running, S. W.: Improvements of the MODIS terrestrial gross and net primary production global data set, *Remote sensing of Environment*, 95, 164–176, 2005.
- 925 Zhu, W., Xie, Z., Zhao, C., Zheng, Z., Qiao, K., Peng, D., and Fu, Y. H.: Remote sensing of terrestrial gross primary productivity: a review of advances in theoretical foundation, key parameters and methods, *GIScience & Remote Sensing*, 61, 2318–2346, 2024.
- Zotta, R., Moesinger, L., van der Schalie, R., Vreugdenhil, M., Preimesberger, W., Frederikse, T., De Jeu, R., and Dorigo, W.: VODCA v2: Multi-sensor, multi-frequency vegetation optical depth data for long-term canopy dynamics and biomass monitoring, <https://doi.org/10.48436/t74ty-tcx62>, data set, 2024a.
- 930 Zotta, R.-M., Moesinger, L., Van Der Schalie, R., Vreugdenhil, M., Preimesberger, W., Frederikse, T., De Jeu, R., and Dorigo, W.: VODCA v2: Multi-sensor, multi-frequency vegetation optical depth data for long-term canopy dynamics and biomass monitoring, *Earth System Science Data*, 16, 4573–4617, 2024b.