Gross primary productivity and the predictability of CO$_2$: more uncertainty in what we predict than how well we predict it

István Dunkl[1,2]**, Nicole Lovenduski**[3]**, Alessio Collalti**[4]**, Vivek K. Arora**[5]**, Tatiana Ilyina**[1]**, and Victor Brovkin**[1,6]

[1]Max Planck Institute for Meteorology, Hamburg, Germany

[2]International Max Planck Research School on Earth System Modelling, Max Planck Institute for Meteorology, Hamburg, Germany

[3]Department of Atmospheric and Oceanic Sciences and Institute of Arctic and Alpine Research, University of Colorado, Boulder, CO, USA CEI

[4]Forest Modelling Laboratory, Institute for Agriculture and Forestry Systems in the Mediterranean, National Research Council of Italy (CNR-ISAFOM), Perugia, Italy

[5]Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, University of Victoria, Victoria, British Columbia, Canada

[6]Center for Earth System Research and Sustainability, University of Hamburg, Hamburg, Germany

Correspondence: István Dunkl (istvan.dunkl@mpimet.mpg.de)

Abstract. The prediction of atmospheric CO$_2$ concentrations is limited by the high interannual variability (IAV) in terrestrial gross primary productivity (GPP). However, there are large uncertainties in the drivers of GPP IAV among Earth system models (ESMs). Here, we evaluate the impact of these uncertainties on the predictability of atmospheric CO$_2$ in six ESMs. We use regression analysis to determine the role of environmental drivers in (i) the patterns of GPP IAV and (ii) the predictability of GPP. There are large uncertainties in the spatial distribution of GPP IAV. Although all ESMs agree on the high IAV in the tropics, several ESMs have unique hotspots of GPP IAV. The main driver of GPP IAV is temperature in the ESMs using the Community Land Model, whereas it is soil moisture in the ESM developed by the Institute Pierre Simon Laplace (IPSL-CM6A-LR) and in the low-resolution configuration of the Max Planck Earth System Model (MPI-ESM-LR), revealing underlying differences in the source of GPP IAV among ESMs. Between 13 % and 24 % of the GPP IAV is predictable 1 year ahead, with four out of six ESMs showing values of between 19 % and 24 %. Up to 32 % of the GPP IAV induced by soil moisture is predictable, whereas only 7 % to 13 % of the GPP IAV induced by radiation is predictable. The results show that, while ESMs are fairly similar in their ability to predict their own carbon flux variability, these predicted contributions to the atmospheric CO$_2$ variability originate from different regions and are caused by different drivers. A higher coherence in atmospheric CO$_2$ predictability could be achieved by reducing uncertainties in the GPP sensitivity to soil moisture and by accurate observational products for GPP IAV.

1 Introduction

Near-term predictions of atmospheric CO$_2$ concentrations are an essential step towards the evaluation of climate mitigation efforts and the development of carbon monitoring programmes (Ilyina et al., 2021). However, the high interannual variability (IAV) in land–atmosphere carbon fluxes, specifically gross primary productivity (GPP), drives the variability in atmospheric CO$_2$ and limits its predictability (Piao et al., 2020). Therefore, the skilful prediction of GPP is a crucial step towards the real-time verification of anthropogenic carbon emissions and the evaluation of mitigation efforts.

The usual approach to evaluate the predictability of an Earth system variable is to compare predictions to observed values. In the case of GPP, this is complicated by the uncertainty in GPP observations (Zhang and Ye, 2021). As

an alternative to calculating the actual predictability that is based on observations, the potential predictability can be assessed by evaluating how well the models can predict their own carbon flux variability CE2. To do this, an ensemble of simulations with an Earth system model (ESM) is initialized from quasi-identical conditions. In a system with little predictability, the spread among the ensemble members increases quickly until all predictive capability is lost when the ensemble spread reaches the magnitude of the IAV. There are, however, certain processes in the Earth system that provide predictability and hinder the divergence of the ensemble members. For example, the El Niño–Southern Oscillation (ENSO) produces predictable climate anomalies that have a sustained impact on GPP (Zeng et al., 2008; Betts et al., 2016). Other processes extend predictability by providing "memory" that maintains the initial conditions. Soils, for example, store initial moisture anomalies by acting as a buffer between the atmosphere and the vegetation (Bellucci et al., 2015). Soil moisture anomalies are further extended through land–atmosphere coupling, which creates a feedback loop that enhances the persistence of these anomalies (Kumar et al., 2020). The initial conditions of the simulations are maintained through the lagged response of plant growth to climatic conditions. Slowly reacting vegetation can cause precipitation anomalies or prolonged drought (Alessandri and Navarra, 2008; Zhang et al., 2021). Given all of these mechanisms of predictability, we find that terrestrial carbon fluxes are predictable for 2 years (Ilyina et al., 2021).

Although several ESMs reproduce the same predictability horizon for globally integrated terrestrial carbon fluxes (Séférian et al., 2018; Ilyina et al., 2021; Spring and Ilyina, 2020; Lovenduski et al., 2019), there are substantial differences in the spatial patterns of GPP IAV (Anav et al., 2015; O'Sullivan et al., 2020). The reason for these differences lies in poorly constrained ecosystem processes that have a large impact on GPP. One of these differences arises from the uncertainty in the sensitivity of GPP to environmental drivers (Ahlström et al., 2015; Jung et al., 2017; Beer et al., 2010; Piao et al., 2020; Collalti et al., 2020). The sensitivity of GPP to temperature and precipitation varies among studies, leading to ongoing discussion concerning the dominant driver of global carbon fluxes (Piao et al., 2020). The different sensitivity of GPP to precipitation across ESMs is further exacerbated by the large disagreement in water storage anomalies (Wu et al., 2021). The simulated annual cycle of water storage anomalies of major river basins is between 0.1 and 2 times that of the observed variability. These deviations in hydrological variability between models are likely to cause similar deviations in GPP IAV, especially in semi-arid watersheds. Further differences in GPP IAV are due to variations in ecosystem boundaries and the related spatial distribution of plant productivity. The Amazon rainforest, for instance, is a hotspot of land–atmosphere carbon fluxes and provides a large contribution to the predictability of atmospheric CO$_2$ (Zeng et al., 2008; Séférian et al., 2018; Ilyina et al., 2021).

However, the transition zone between the wet tropical forest and semi-arid tropics within the Amazon Basin varies among the models due to differences in their representation of land cover (Collier et al., 2018; Hu et al., 2022). Such differences in biome boundaries also modify the impact of ENSO on GPP IAV. ENSO produces distinct spatial patterns of climatic anomalies that significantly influence the GPP on 32 % of the vegetated land area (Zhang et al., 2019). These ENSO-related climate patterns will have a different impact on GPP depending on the type of biomes under their influence. In addition to the spatial variability, many ESMs struggle to reproduce the seasonal variability in carbon fluxes. This can be seen in the large biases in phenology (Song et al., 2021). Several models overestimate the seasonal amplitude of leaf area index (LAI) in the tropics and mismatch the timing of the LAI maxima and minima (Peano et al., 2019, 2021).

All of these uncertainties suggest that there are large differences in the patterns of GPP IAV among the ESMs, but it is currently unclear how these differences affect the predictability of GPP. With this study, we want to extend our understanding of GPP predictability by considering the different patterns of GPP IAV among the ESMs. In a multimodel analysis, we investigate which processes drive the IAV in GPP and which processes allow the GPP IAV to be predictable. Regression analysis is used to determine the role of three environmental variables (soil moisture, temperature, and radiation) on GPP IAV and GPP predictability. We analyse the cause of differences in GPP predictability across ESMs, identify the areas of large discrepancies, and determine the factors contributing to the attached uncertainties. The aim of this study is to reveal which factors of GPP representation are limiting the predictability of atmospheric CO$_2$.

2 Methods

2.1 Data sources

We analyse model output from the Decadal Climate Prediction Project (DCPP; Boer et al., 2016). This protocol-driven multi-model approach aims at studying the decadal predictability of the Earth system with hindcasts, quasi-real-time forecasts, and case studies on predictability mechanisms. The hindcasts are initialized annually from 1960 to 2017 or 2019 with the starting dates between November and January and at least 10 ensemble members. Simulations are driven by Coupled Model Intercomparison Project Phase 5 (CMIP5) or Phase 6 (CMIP6) historical forcing and extended by Representative Concentration Pathway (RCP) 4.5 or Shared Socioeconomic Pathway (SSP) 2-4.5 afterwards. The DCPP framework does not prescribe any specific initialization or data assimilation methods and leaves these details to be decided by the respective modelling centres.

We additionally use the Community Earth System Model 2 (CESM2) output from the Seasonal-to-Multiyear Large En-

semble (SMYLE) prediction system (Yeager et al., 2022). The SMYLE hindcasts ensembles are initialized four times per year with 20 ensemble members between 1970 and 2019. In this study, the November initializations are used to achieve the highest comparability with the DCPP hindcasts.

We compare the spatial GPP IAV patterns of the ESMs with observation-based GPP products. Because of the uncertainty among observations, we include products based on three different sources. The Vegetation Photosynthesis Model (VPM; Zhang et al., 2017) is a remote-sensing-based product that uses a light use efficiency (LUE) model to calculate GPP. VPM uses satellite data from MODIS and an improved LUE algorithm that considers leaf quality. The second data set is GOSIF (Li and Xiao, 2019) which is based on data from MODIS and the Orbiting Carbon Observatory-2. GOSIF uses solar-induced chlorophyll fluorescence, which is a more recent approach, to calculate GPP. Lastly, we use FLUXCOM (version RS–METEO, ERA5; Jung et al., 2019), which uses machine learning to upscale flux tower observations with meteorological and remote-sensing data. Because FLUXCOM underestimates the IAV in GPP (Anav et al., 2015; O'Sullivan et al., 2020), it is recommended to scale the data so that the IAV of integrated FLUXCOM fluxes resembles observations (Jung et al., 2019). The VPM, GOSIF, and FLUXCOM data are linearly detrended before calculating the IAV. Due to its long time span, FLUXCOM is detrended over two periods (1979–1999 and 2000–2018).

CanESM5

The Canadian Earth System Model version 5 (CanESM5; Swart et al., 2019) consists of the Canadian Land Surface Scheme (CLASS) and Canadian Terrestrial Ecosystem Model (CTEM) with a T63 grid with an approximate resolution of 2.8°. The atmosphere is realized with the Canadian Atmospheric Model (CanAM5) with 49 vertical levels. Ocean physics is simulated with CanNEMO, on a tripolar grid with a resolution of 1 to 1/3° and 45 vertical levels, and ocean biogeochemistry is represented by the Canadian Model of Ocean Carbon (CMOC).

The CanESM5 hindcast simulations are initialized every January between 1960 and 2017 with 20 members. The 3D potential temperature and salinity of the global oceans are nudged toward the monthly Ocean Reanalysis System 5 (ORAS5; Zuo et al., 2019). Sea surface temperatures are nudged to the Extended Reconstructed Sea Surface Temperature (ERSSTv3; Xue et al., 2003; Smith et al., 2008) until 1981 and to the Optimum Interpolation Sea Surface Temperature (OISST; Banzon et al., 2016) afterwards. Sea ice concentration is nudged to the Hadley Centre Sea Ice and Sea Surface Temperature data set (HadISST.2; Titchner and Rayner, 2014), and sea ice thickness is nudged to monthly climatology until 1988 and to the SMv3 statistical model of Dirkson et al. (2017) afterwards. For the atmosphere, temperature, horizontal wind components, and specific humidity

are nudged to ERA-40 (Uppala et al., 2005) until 1978 and to 6-hourly ERA-Interim data (Dee et al., 2011) afterwards.

CESM1-CAM5

The Community Earth System Model (CESM) version 1.1 (Hurrell et al., 2013) is used to produce 40-member simulations in the Decadal Prediction Large Ensemble (DPLE) project (Yeager et al., 2018). The model components are the Community Land Model version 4 (CLM4; Lawrence et al., 2011) with a 1° resolution, the Community Atmosphere Model version 5 (CAM5) with 30 vertical levels, the Parallel Ocean Program (POP2) with 60 vertical levels, and sea ice with the Community Ice Code (CICE4).

The CESM1-CAM5 hindcasts are initialized every November. There is no direct assimilation of observations to produce the initial conditions; instead, ocean and sea ice are obtained from simulation runs forced by historic atmospheric surface fields (Yeager et al., 2018). Initial conditions for the land and atmosphere components are obtained from ensemble member no.34 of the CESM Large Ensemble (Kay et al., 2015; Lovenduski et al., 2019).

CESM2

CESM version 2 (Danabasoglu et al., 2020) runs on a 1° horizontal resolution for all components. The atmosphere is simulated by the Community Atmosphere Model version 6 (CAM6) with 32 vertical levels. The ocean model is the Parallel Ocean Program version 2 (POP2) with 60 vertical levels, with the biogeochemistry from the Marine Biogeochemistry Library and sea ice by CICE version 5.1.2 (CICE5) with 8 vertical layers. The land component is simulated by the Community Land Model version 5 (CLM5; Lawrence et al., 2019), which has several updates to its predecessors CLM4 and CLM4.5, leading to a better representation of the global carbon cycle in benchmarks (Bonan et al., 2019).

Hindcasts are initialized on the first day of every November, February, May, and August, and they run for 24 months. Only the November initializations are used in this analysis to increase comparability with the DCPP simulations. Initial conditions for the atmosphere, ocean, and sea ice stem from the Japanese 55-year Reanalysis – JRA-55 (Kobayashi et al., 2015) and JRA55-do (Tsujino et al., 2018). The land surface and biogeochemistry are initialized from forced CLM5 simulations.

CMCC-CM2-SR5

The Euro-Mediterranean Centre on Climate Change coupled climate model (CMCC-CM2; Cherchi et al., 2019; Lovato et al., 2022) is based on CESM and consists of the Community Land Model (CLM4.5) with a 1° resolution and the atmospheric model CAM5.3 with 30 vertical levels. The distinguishing element of CMCC-CM is the ocean, which is simulated by NEMO3.6, while sea ice is modelled by CICE4.
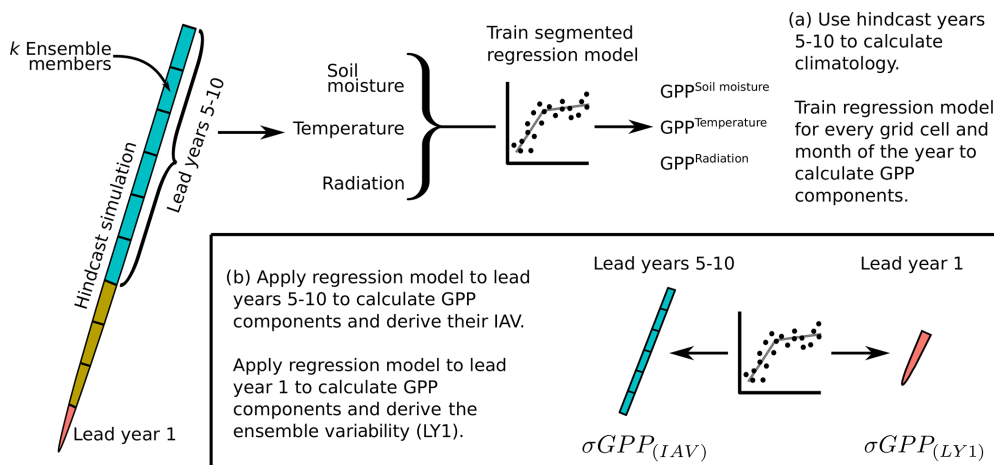
Figure 1. Workflow of the statistical analysis. In panel **(a)**, lead years 5 to 10 of the hindcast simulations are used to train a regression model that calculates the components of GPP caused by the environmental drivers. In panel **(b)**, the regression model is applied to lead years 5 to 10 to calculate the IAV in the GPP components ($\sigma\mathrm{GPP}_{(\mathrm{IAV})}$) and to lead year 1 to calculate the mean ensemble spread of the GPP components ($\sigma\mathrm{GPP}_{(\mathrm{LY1})}$).

The 10-member hindcast simulations are initialized every November (Nicolì et al., 2023). The ocean initial conditions are from CHOR (Yang et al., 2017) until 2010 and from CGLORSv7 (Storto and Masina, 2016) afterwards. The atmosphere is initialized from ERA-40 until 1978 and from ERA-Interim afterwards. The land surface is initialized using the reanalysis with two different meteorological forcings. For this reason, only ensemble members 1, 3, 5, 7, 8, and 9 are used, as the other members start from a different state and this would not allow for the quantification of predictability by ensemble spread.

Because the CMCC-CM2-SR5 fields containing land–atmosphere carbon fluxes are not exported for the DCPP runs, the historical simulations are used to infer the relationship between environmental drivers and GPP.

IPSL-CM6A-LR

The ESM developed by the Institute Pierre Simon Laplace (IPSL; Boucher et al., 2020) uses the ORCHIDEE v2.0 (Cheruy et al., 2020) land surface model (LSM) with an average resolution of 157 km. The atmosphere is simulated at the same resolution by LMDZ6 with 79 vertical levels, the ocean is simulated with NEMO-OPA with a 1° resolution and 75 vertical levels, and ocean biogeochemistry is simulated with PISCESv2.

The hindcast simulations of IPSL-CM6A-LR come from the DCPP project. The 10-member ensembles start annually in January between 1960 and 2016. The hindcasts are initiated from an assimilation run with EN4 sea surface temperatures (Good et al., 2013) and Atlantic sea surface salinity (Estella-Perez et al., 2020). Subsurface ocean, sea ice, and atmosphere are not assimilated.

MPI-ESM-LR

MPI-ESM-LR is the Max Planck Earth System Model (MPI-ESM1.1; Giorgetta et al., 2013) used in a low-resolution configuration. The land is simulated by JSBACH with dynamic vegetation (Reick et al., 2013). The ocean component is the Max Planck Institute for Meteorology Ocean Model (MPIOM) with a horizontal resolution of about 150 km and 40 vertical levels. The atmosphere is simulated by ECHAM at a T63 resolution with 47 vertical layers, and ocean biogeochemistry is represented by the Hamburg Ocean Carbon Cycle (HAMOCC) model.

The utilized hindcast simulations of MPI-ESM-LR are conducted within the MiKlip project (Marotzke et al., 2016). The decadal prediction system comprises 10-member ensembles starting every January between 1961 and 2014. Ocean temperature and salinity are initialized from the Ocean Reanalysis System 4 (ORAS4; Balmaseda et al., 2013), and the atmosphere is initialized from ERA-40 from 1960 to 1998 and from ERA-Interim from 1990 to 2014.

2.2 Statistical approach

Overview

An overview of the statistical analysis is shown in Fig. 1. Every hindcast simulation is initialized from quasi-identical conditions. With the increasing lead time, the variability within the hindcast ensemble (standard deviation across the ensemble members for a given time) also increases until it reaches the IAV. Based on this assumption, the hindcast simulations are split into two groups by lead time: lead year 1 and lead years 5 to 10. For lead years 5 to 10, the effects of ocean and atmosphere initialization are assumed to be neg-

ligible. These years are used to calculate the monthly mean climatology, which is removed from both groups to obtain the anomalies. The anomalies of lead years 5 to 10 are used in a regression analysis to derive the sensitivity of GPP to the environmental drivers, i.e. soil moisture, temperature, and radiation (Fig. 1a). The regression model is applied to the anomalies of lead years 5 to 10 to calculate the IAV in all GPP components ($\sigma\text{GPP}_{\text{(IAV)}}$) and to the anomalies of lead year 1 to calculate the ensemble variability in all GPP components ($\sigma\text{GPP}_{\text{(LY1)}}$) (Fig. 1b). We derive the predictability of GPP by comparing $\sigma\text{GPP}_{\text{(LY1)}}$ to $\sigma\text{GPP}_{\text{(IAV)}}$. Because the hindcast simulations are not evaluated against observations, the calculated predictability reflects the potential predictability.

Climatology and sensitivity

The monthly mean climatologies are calculated from lead years 5 to 10, with a 3-year moving-window approach for every calendar year. Because the moving-window method is not applicable for the first decade of hindcast initializations, the monthly climatology for the 1960s (1970s for CESM2) is calculated based on all lead years 5 to 10 within the 1960s (or 1970s). Anomalies of all input fields are calculated by subtracting the monthly climatologies from the hindcast data. The obtained anomalies of lead years 5 to 10 make up a data set of n simulation years:

$$n = 6 \text{ hindcast years} \times \text{no. ensemble members}$$
$$\times \text{ no. initializations.} \tag{1}$$

A total of 10 to 40 ensemble members and 56 to 58 initializations result in sample sizes of 3330 to 13 680. Because the hindcast length is only 2 years for the CESM2 simulations, a different approach is used here. Instead of lead years 5 to 10, only lead year 2 is selected and only five random ensemble members are used from every hindcast to reduce the number of simulations with the same initial conditions. To offset the reduced number of data points, five random simulations are also added from the hindcast simulations initialized in February, May, and August.

The resulting data set of lead year 5 to 10 anomalies is used to derive the sensitivity of GPP to the environmental drivers (ENV: soil moisture, temperature, and radiation) by fitting a regression model for every grid cell and month of the year. The relationship between GPP and the environmental drivers is frequently non-linear, sometimes due to specific breakpoints in the functional representation of GPP. For this reason, segmented linear regression (SLR) is used to model GPP from the environmental drivers (Muggeo, 2008). SLR finds breakpoints in the data, splitting them into multiple ranges and fitting an individual regression model to each of the data ranges. Here, a single breakpoint is determined for each of the three predictor variables.

Because environmental drivers have some degree of collinearity, the regression analysis will not be able to fully attribute the GPP anomalies to their specific causes. Therefore, the resulting sensitivities should be taken as a "contributive" and not a "true" effect of the environmental drivers (Wang et al., 2016).

Variability and predictability

The SLR can now be applied to individual simulations to determine the component of GPP anomalies that can be attributed to each of the environmental drivers:

$$\Delta\text{GPP} \approx \Delta\text{GPP}^{\text{Soil moisture}} + \Delta\text{GPP}^{\text{Temerature}}$$
$$+ \Delta\text{GPP}^{\text{Radiation}}. \tag{2}$$

The three components of GPP anomalies ($\Delta\text{GPP}^{\text{ENV}}$) are calculated for every simulation within the hindcast lead time 5 to 10. From the results, we calculate the IAV in the components for every grid cell and month of the year ($\sigma\text{GPP}_{\text{(IAV)}}^{\text{ENV}}$). Similarly, the SLR is applied to the anomalies of lead year 1 to calculate the standard deviation for every month within the hindcast simulations. Averaging over the standard deviations of every hindcast returns the ensemble variability in lead year 1 ($\sigma\text{GPP}_{\text{(LY1)}}^{\text{ENV}}$).

The predictability is assessed by comparing $\sigma\text{GPP}_{\text{(LY1)}}$ to $\sigma\text{GPP}_{\text{(IAV)}}$. A high predictability of an input field means that the ensemble variability is restricted for some time after the hindcast initialization and does not reach the IAV immediately (Fig. 2a). In this study, we use two metrics to evaluate different aspects of predictability. We calculate the fraction of GPP IAV that is predictable (the predictable fraction – pf) to assess the ability of a system to retain memory. Although this metric is useful for quantifying the mechanisms that provide predictability at a local level, the pf is not suitable for assessing how GPP predictability affects the predictability of atmospheric CO$_2$. This is because the regions with a high pf do not necessarily contribute much to the global GPP fluxes. The regions with the highest pf values are often in deserts with very low carbon fluxes (Dunkl et al., 2021). To assess the contribution of GPP predictability to atmospheric CO$_2$ predictability, we calculate the absolute portion of the IAV that can be predicted as the predictable component (pc). The pc is the difference between IAV and ensemble variability and is generally higher in regions that contribute more to CO$_2$ IAV:

$$\text{pc}^{\text{ENV}} = \sigma\text{GPP}_{\text{(IAV)}}^{\text{ENV}} - \sigma\text{GPP}_{\text{(LY1)}}^{\text{ENV}}, \tag{3}$$

$$\text{pf}^{\text{ENV}} = \frac{\text{pc}^{\text{ENV}}}{\sigma\text{GPP}_{\text{(IAV)}}^{\text{ENV}}}. \tag{4}$$

The use of the two predictability metrics is exemplified in Fig. 2b.
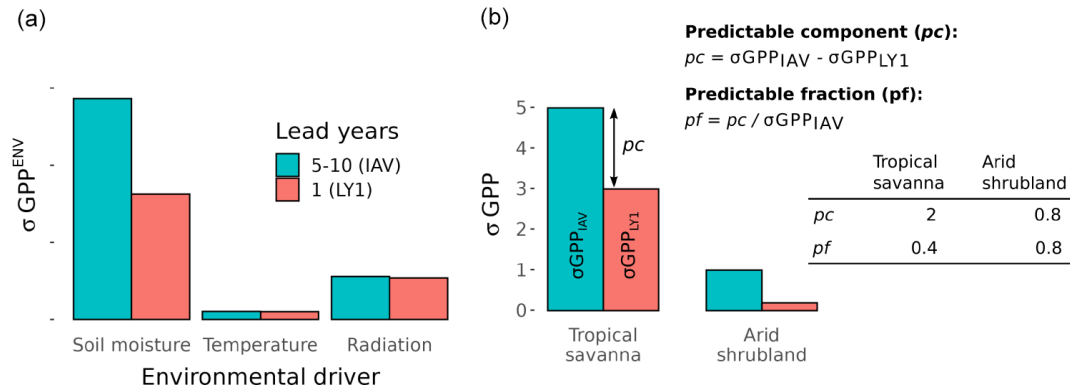
Figure 2. Panel (**a**) presents the exemplary composition of GPP variability and predictability in a tropical forest. The components of GPP IAV are calculated from lead years 5 to 10 (green bars) and the ensemble variability is calculated from lead year 1 (red bars). In the exemplified region, most of the variability is caused by soil moisture and radiation, and GPP is not restricted by temperature. Predictability is exclusively provided through soil moisture. Panel (**b**) demonstrates the need for the two predictability metrics using a tropical savanna and an arid shrubland as examples. The predictable component (pc) is the absolute predictable IAV, and the predictable fraction (pf) is the pc scaled by the IAV. While the arid shrubland has a better potential to retain memory (as seen from the high pf), these ecosystems contribute little to the variability in atmospheric CO$_2$, which can be better assessed using the pc.

3 Results and discussion

3.1 Patterns of GPP IAV

In order to understand what the models are predicting, we start by analysing the patterns of GPP IAV. There are differ-
5 ences in the overall magnitude of GPP IAV among ESMs, with CanESM5, CMCC-CM2-SR5, and IPSL-CM6A-LR at the lower end of the IAV spectrum, whereas CESM2 and MPI-ESM-LR are at the higher end of the IAV spectrum. Factors that could explain some of the differences in the over-
10 all magnitude of IAV are the relatively weak ENSO telecon-nection in CanESM5 (Swart et al., 2019) or the low total GPP in CMCC-CM2-SR5 (Lovato et al., 2022).

Because we focus on the spatial patterns of IAV rather than absolute differences, the GPP IAV patterns are scaled
15 for better comparison (Fig. 3). We find agreement in the large-scale patterns of GPP IAV, with most of the IAV in the ESMs in the northern Amazon Basin and in the semi-arid tropics like western South America, southern Africa, South Asia, Australia, and southern North America (detailed
20 maps of the location of the semi-arid regions in the ESMs are shown in Fig. 5). A closer examination of GPP IAV re-veals that the ESMs show less agreement in the regions con-tributing most to the IAV, especially in the semi-arid trop-ics. Some ESMs have large hotspots of GPP IAV that can-
25 not be found in other ESMs. These unique hotspots are the western Amazon Basin (CanESM5), central South America (CESM2), southern Africa (MPI-ESM-LR and CanESM5), and Australia (MPI-ESM-LR). We find the most consistency on the northern coast of South America, which is a high-IAV
30 region in most ESMs. The spatial patterns of IAV have an average correlation of 0.47 among the ESMs. The ESM with the lowest correlation values is CESM2, with an average of

0.29. CESM2 stands out with a very low IAV in the tropical rainforests of the Amazon and Congo basins and in Southeast Asia. 35

The correlation among the observational products is 0.65; although the products CE3 confirm most of the IAV patterns of the ESMs, we find stronger deviations in South Amer-ica. While many ESMs have IAV hotspots along the northern coast of South America, this is only reproduced in FLUX- 40 COM. However, all observational products show a high GPP IAV in western South America, which can not be found in the ESMs. The spatial patterns of GPP IAV revealed here correspond to the literature, which suggests that the semi-arid tropics, tropical forests, grasslands, and croplands are 45 the main drivers of global GPP IAV (Ahlström et al., 2015; Piao et al., 2020; O'Sullivan et al., 2020). These studies also reflect the large uncertainty in the contribution of the indi-vidual semi-arid regions to GPP IAV between the models, in particular the uncertain role of Australia. In an ensemble 50 of eight LSMs, Australia contributed 39 %, semi-arid tropi-cal Africa contributed 32 %, and Southeast Asia contributed 10 % to global GPP IAV, whereas temperate South America only contributed 2 % (Chen et al., 2017). Although Australia has the highest mean model IAV, the variability in IAV be- 55 tween the models is also the largest, with GPP IAV ranging between 0.26 and 1.01 PgC yr^{-1}. While the large role of the dry tropics in driving GPP IAV is not disputed, it is likely that ESMs underestimate GPP IAV in wet tropical forests (O'Sullivan et al., 2020). This results from the limited avail- 60 ability of observations due to few flux towers and from the fact that the quality of remote-sensing products is limited in tropical forests due to saturation effects and a high cloud cover (Kolby Smith et al., 2016). In this study, the low GPP
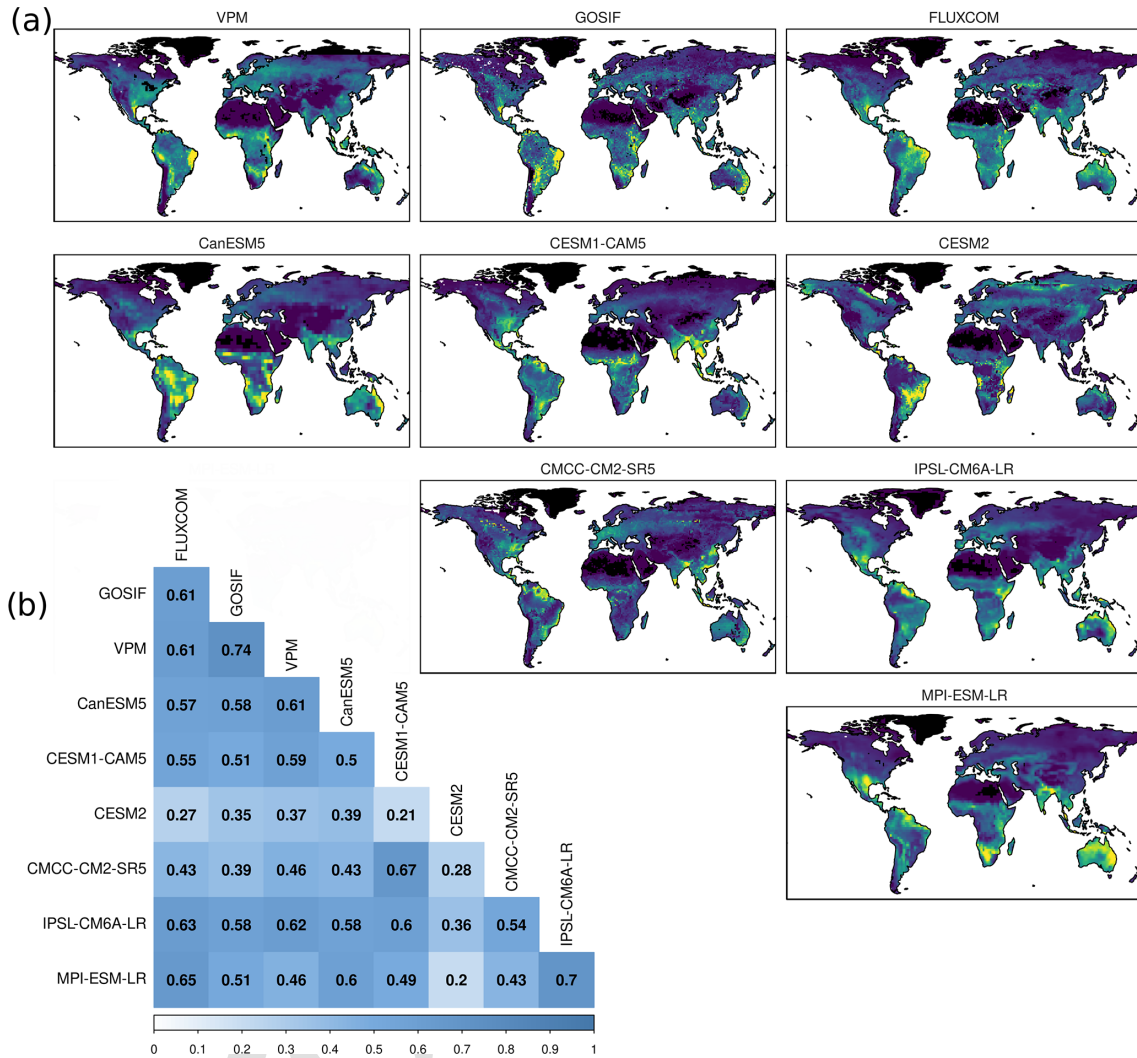
Figure 3. GPP IAV in three observational products (VPM, GOSIF, and FLUXCOM) and six ESMs. Panel **(a)** presents the spatial patterns of GPP IAV, with brighter colours representing higher values. The data are scaled across ESMs to highlight differences in patterns and not absolute differences. Panel **(b)** displays the spatial correlations between the products.

IAV in tropical forests is especially evident for CESM2, as IAV increases abruptly outside the wet tropical forests.

The divergence in GPP IAV across different ESMs is largely caused by three factors: the sensitivity of carbon fluxes to climatic drivers (Piao et al., 2020), as discussed in Sect. 3.2; phenology (Chen et al., 2017; Peano et al., 2019, 2021); and meteorological input (Anav et al., 2015). The role of phenology is crucial because the amount and quality of leaves determine the carbon fluxes between land and the atmosphere (Peano et al., 2021). Most LSMs tend to have a better representation of the growing season type and growing season boundaries in the wet than in the semi-arid tropics. Peano et al. (2021) analysed the start and end months of growing seasons in eight LSMs under the same climate forcing and found several regions with a wide range of simulated growing seasons. The largest uncer-

tainties in the growing season are in the semi-arid tropics, the same regions in which we find little agreement in GPP IAV. The start of the growing season ranges from February to October in Australia and from March to October in southern Africa, while the end of the growing season ranges from March to September in Africa between 0 and 15° N. The vegetation types with the largest uncertainty in growing season timing are broadleaf, deciduous shrubs, which are mostly located in northern Australia; Southern Hemisphere crops; broadleaf, evergreen trees; and grasses. The better-performing LSMs have a high number of plant functional types or use more complex phenology schemes that depend on plant functional types, and they also use a larger number of environmental variables to constrain phenology. Its complex phenological scheme puts ORCHIDEE among the better-performing LSMs and might also explain the high cor-
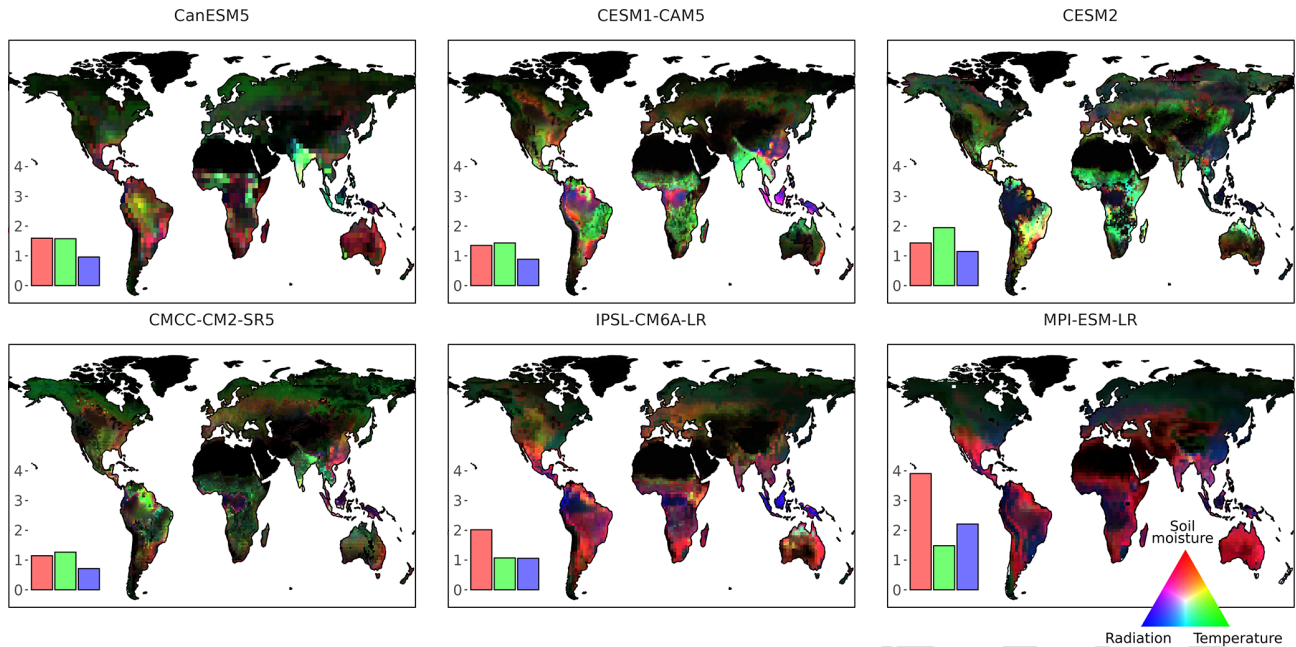
Figure 4. The contribution of environmental drivers to GPP variability ($\sigma\,\mathrm{GPP}^{\mathrm{ENV}}_{\mathrm{(IAV)}}$). Colour intensity represents higher GPP IAV. The data are scaled across ESMs to highlight differences in patterns, not absolute differences. Bars represent the mean contribution of environmental drivers to global GPP IAV ($\mathrm{kg\,C\,m^{-2}\,s^{-1}\,10^{-13}}$) CE4.

relation of IPSL-CM6A-LR with the GPP IAV for all three observational products.

A misrepresentation of phenology could also explain the overall high IAV in MPI-ESM-LR. JSBACH overestimates the seasonality of the LAI in the tropics, and this becomes visible in the strong seasonal cycle of tropical LAI in MPI-ESM-LR (Song et al., 2021). Consequentially, the area of the evergreen tropics is underestimated in JSBACH (Peano et al., 2021). This leads to a larger fraction of semi-arid tropics with a higher GPP IAV. This amplification of the equatorial dry season might lead to the high GPP IAV in the northern Amazon and contribute to the overall high IAV in MPI-ESM-LR (Wang et al., 2011).

3.2 Drivers of GPP IAV

To determine the drivers of GPP IAV, we analysed the sensitivity of GPP to environmental drivers using regression analysis. The globally averaged contribution of the drivers to GPP IAV is shown as the bars in Fig. 4. The CLM family and CanESM5 have similar patterns, with temperature dominating the IAV or being on par with soil moisture. IPSL-CM6A-LR and MPI-EMS-LR have distinctly different patterns, with soil moisture dominating the IAV and radiation contributing equally or more than temperature. A reason for the large contribution of soil moisture to GPP IAV in IPSL-CM6A-LR and MPI-ESM-LR could be that both ESMs are at the high end of soil moisture IAV for deep soil layers in the Southern Hemisphere (Qiao et al., 2022), where many

of the semi-arid ecosystems are located that contribute most to GPP IAV. Another explanation could be that out of 11 ESMs, IPSL-CM6A-LR and MPI-ESM-LR have the lowest warm-season soil moisture (Padrón et al., 2022). This increase in dryness can lead to a larger extent of semi-arid ecosystems with a generally higher GPP IAV. Another effect of reduced warm-season soil moisture can be an increase in the land–atmosphere coupling strength (Santanello et al., 2018). Stronger land–atmosphere coupling would explain the higher correlation between soil moisture and temperature in IPSL-CM6A-LR and MPI-ESM-LR (Padrón et al., 2022) and would make the regression coefficients shift towards the stronger predictor, which is soil moisture.

The spatial drivers of GPP IAV show agreement in the wet and arid tropics, whereas there is little consistency in the semi-arid transition zones (Fig. 4). In many ESMs, the GPP IAV in the wet tropics and in eastern China is induced by radiation, while soil moisture becomes more prevalent along the aridity gradient and is driving IAV in southern Africa, southern South America, and Australia. The IAV in the remaining land surface is driven predominantly by soil moisture in IPSL-CM6A-LR and MPI-ESM-LR, whereas it is driven by a combination of temperature and soil moisture in the remaining ESMs.

Some of the differences in GPP sensitivity among the ESMs can be explained by differences in aridity. A higher sensitivity to soil moisture can result from a dryer climate. The distribution of climate zones in the analysed models is shown in Fig. 5 using the De Martonne aridity index
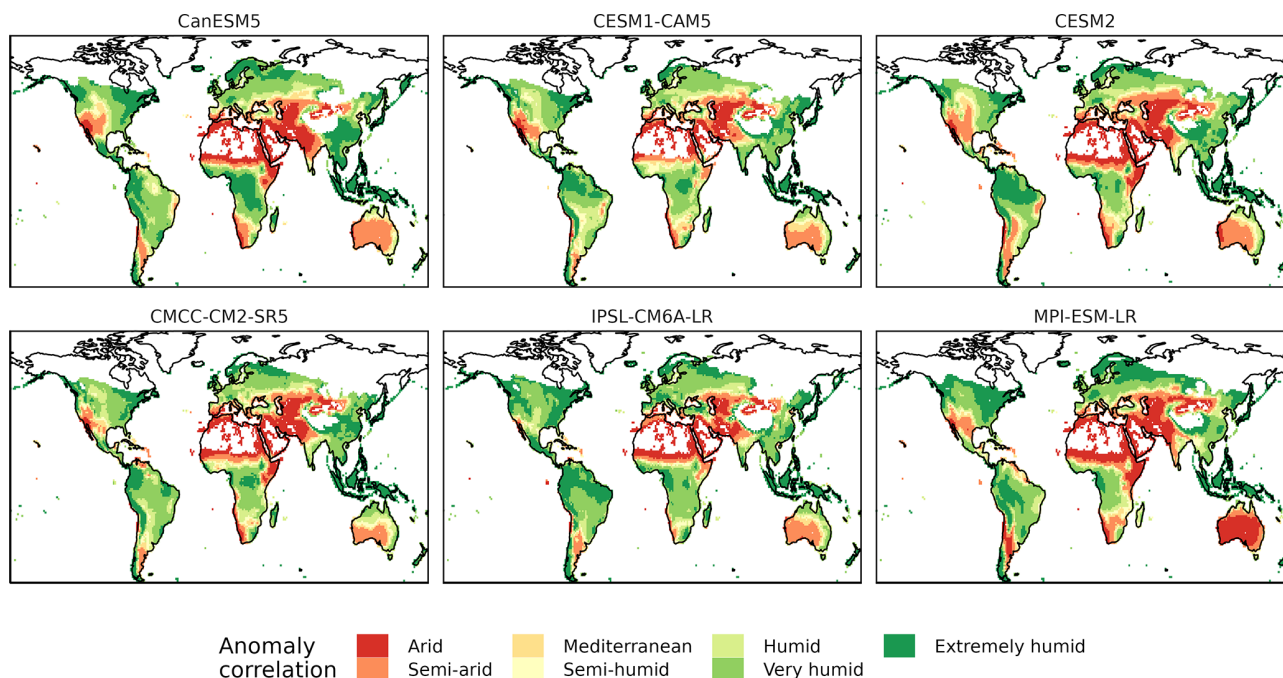
Figure 5. De Martonne aridity index of the analysed models.

(Gavrilov et al., 2019). MPI-ESM-LR and CanESM5 show an above-average extent of arid and semi-arid regions in Australia and southern Africa. This could explain the high sensitivity of GPP to soil moisture in these regions. Differences in climate also explain some of the discovered GPP patterns in the Amazon Basin. CESM2 is the model with the most humid climate in the Amazon Basin, which could be the reason for the low sensitivity of GPP to soil moisture and the generally low GPP IAV in this region. In contrast, we find that CanESM5 is on the other side of the spectrum, with a relatively dry Amazon Basin, leading to a higher sensitivity to soil moisture and a high GPP IAV. However, there are also differences in GPP sensitivity that cannot be explained by differences in climate. IPSL-CM6A-LR is more or equally humid in Australia, southern Africa, and South America than the models of the CLM family, despite having a high sensitivity of GPP to soil moisture in these regions. These differences are more likely to be caused by differences in their land surface models than by climate.

The general patterns of GPP sensitivity agree with reported sensitivity patterns in the literature. Multi-model averages and observations of GPP sensitivity agree with the larger role of temperature in tropical forests, radiation in western Amazonia, and the importance of precipitation in the semi-arid tropics (O'Sullivan et al., 2020; Anav et al., 2015). However, the role of water on carbon fluxes increases when soil moisture is used instead of precipitation in sensitivity studies (Piao et al., 2020). This can be observed in the sensitivity of net biome productivity (NBP), which shows a more balanced contribution of soil moisture and temperature in the

wet tropical forests (Piao et al., 2020; Padrón et al., 2022). Although the comparison of GPP and NBP imposes limitations, GPP explains the majority of tropical NBP (Ahlström et al., 2015). This suggests that the low water sensitivity of tropical GPP might explain the lower than expected GPP IAV in tropical forests in ESMs.

3.3 Predictability of GPP

To analyse the role of GPP in the predictability of atmospheric CO$_2$, we assessed GPP predictability using two metrics. The predictable component (pc) is calculated as the difference between ensemble variability and IAV and provides a measure of absolute predictable variability. The pc can be used to assess the predictability of GPP fluxes that contribute to CO$_2$ variability. The predictable fraction (pf) is the ratio of pc to IAV and illustrates how well memory is retained in the system. This metric can be used to compare the predictive performance of different biomes, for example.

There is relatively high consistency among the pf values of the environmental drivers across the models (pf$^{\text{Soil moisture}}$ > pf$^{\text{Temperature}}$ > pf$^{\text{Radiation}}$; the numbers above the bars in Fig. 6). This pattern reflects the anticipated differences in predictability among the drivers. Atmospheric fields, as radiation, have a low persistence, leading to a low predictability of 2 weeks for most regions (Zeng et al., 2008). Soil hydrology, on the other hand, acts as a low-pass filter that removes the unpredictable high-frequency variability in precipitation and allows a predictability of soil moisture of around 2 years (Chikamoto et al., 2017). Temperature gains most
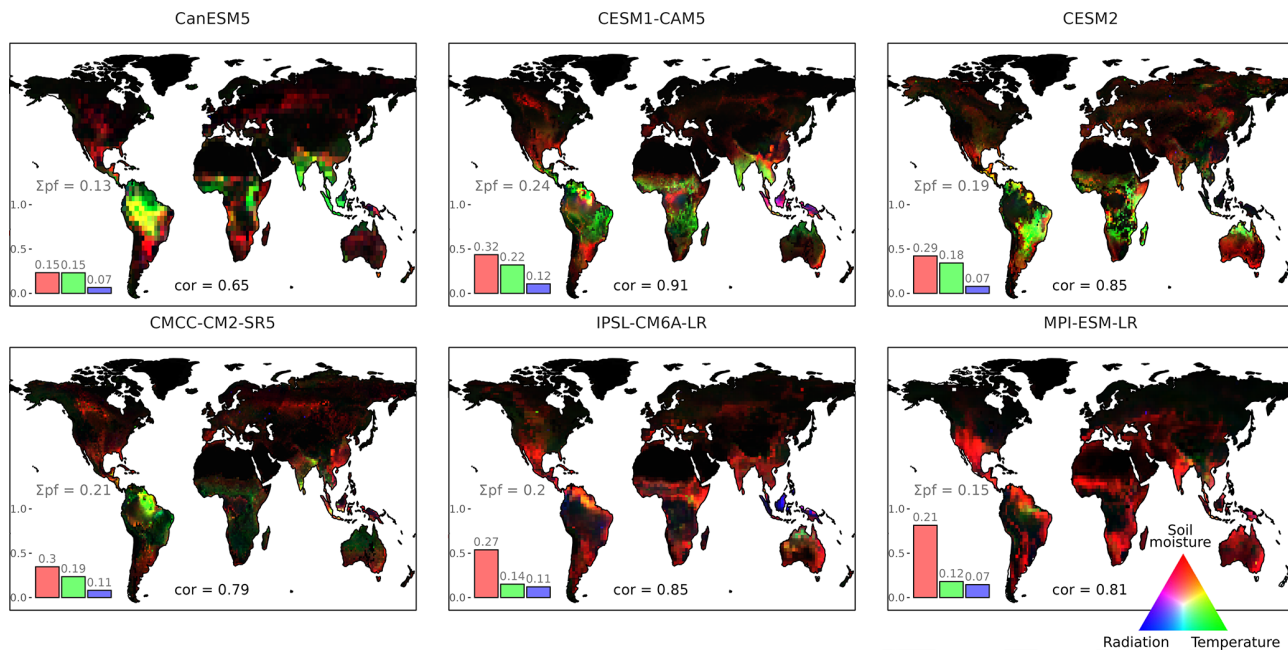
Figure 6. The contribution of environmental drivers to the predictable component (pc) of GPP. The contribution is calculated as the difference between the IAV and the ensemble variability within lead year 1 of the hindcast experiments. Values are scaled for each ESM. Bars represent the mean contribution of environmental drivers to the pc ($\Delta\sigma$ GPP, in kg C m^{-2} s^{-1} 10^{-13}). Numbers on top of the bars show the predictable fraction (pf), which is the share of the pc to overall IAV. The correlation between GPP IAV and the pc is shown at the bottom of the plots.

of its predictability through sea surface temperature (SST) forcing in the equatorial regions (Feng et al., 2011) and via land–atmosphere coupling in the semi-arid tropics (Seo et al., 2019).

The overall pf of CESM2, CMCC-CM2-SR5, and IPSL-CM6A-LR falls into a narrow window of 0.19 to 0.21. CESM1-CAM5 has the highest pf with a value of 0.24. It is likely that this increased share of predictable IAV is not due to differences in model structure but rather due to the large number of ensemble members (40). Most other ESMs in this study have only 10 ensemble members, which is not enough to capture the difference between ensemble variance and IAV; thus, an increase in ensemble members leads to an increase in prediction skill (Meehl et al., 2021). However, despite having 20 ensemble members, CanESM5 has the lowest pf among the models. A possible explanation could be the low IAV in deep-layer soil moisture in CanESM5 (Qiao et al., 2022). A limited ability to reproduce the full spectrum of soil moisture variability could mean that the soils have a smaller buffering capacity. As a result, they are not able to simulate the observed persistence of soil moisture anomalies, leading to a reduction in predictability. On the other hand, a high variability in soil moisture does not guarantee a high pf, as seen, for example, for MPI-ESM-LR. The low pf of MPI-ESM-LR can be explained by the sensitivity of GPP to radiation. As only 7 % to 12 % of the radiation-induced IAV is predictable, a high share of σGPP$^{\text{Radiation}}$ reduces the predictability of GPP. This becomes evident in MPI-ESM-LR,

in which the share of σGPP$^{\text{Radiation}}$ is 20 % higher than in the other ESMs.

We find that the regions contributing to the predictability of atmospheric CO$_2$ (pc) are highly related to the IAV patterns. The spatial correlation between pc and IAV exceeds 0.79 in all models except CanESM5. Indeed, these high correlations between predictability and IAV align with our understanding. Under a constant pf, pc would grow linearly with increasing IAV, leading to a perfect correlation. These high correlations show that the differences in the predictability of atmospheric CO$_2$ are determined more by the differences in GPP IAV than the differences in the pf of GPP. While the pf values show that the ESMs have a similar degree of memory retention, there are few overlaps in the spatial distribution of the pc, with an average correlation of 0.38 between the ESMs. For an alternative quantification of this disagreement, we separate the high-predictability grid cells, which are the grid cells contributing to the top 20th quantile of pc. A total of 74 % of these high-predictability grid cells are unique to only one ESM, and only 8 % of high-predictability grid cells can be found in three or more ESMs.

Although the spatial patterns of the pc broadly resemble the patterns of GPP IAV, there are some slight differences between these fields. The pc is relatively high along the northeastern coast of South America in most ESMs. This could be due to the high climate predictability caused by slowly evolving Atlantic SST patterns (Dirmeyer et al., 2018). Other systematic differences can be explained by the differing pc

of the environmental drivers. The most evident is the difference between IAV and predictability in regions where GPP IAV is driven by radiation. This leads to relatively low predictability in the tropical rainforests of the western Amazon Basin and the Congo Basin. An exception is the predictability provided through radiation on the Southeast Asian islands in IPSL-CM6A-LR and CESM1-CAM5. High predictability in these regions could be explained by the proximity to the ENSO SST region. Strong and predictable SST anomalies in the tropical Pacific that surround the islands can directly influence the cloud cover over land. The predictable component is also higher over areas where IAV is driven by soil moisture rather than temperature. In many ESMs, this leads to a high predictable component in the semi-arid regions of South America, Africa, and India.

4 Conclusions

We tested the ability of six ESMs to predict terrestrial GPP and determined their similarities and the sources of uncertainties. The ESMs are fairly similar in their ability to retain memory in hindcast simulations and predict their own variability, with the pf values of four of the ESMs falling between 19 % and 24 %. Most of the GPP pf is provided by soil moisture. Up to 32 % of the GPP IAV caused by soil moisture is predictable, whereas this value is only 7 % to 12 % for the IAV caused by radiation. The differences in the pf values among ESMs are due to the ensemble size and the sensitivity of GPP to radiation. Further sources of predictability that are not studied here are long-term vegetation dynamics. Specifically, the large and structural changes like tree mortality (Wigneron et al., 2020) and recruitment (Holmgren et al., 2001). These processes only occur in extreme years and cause shifts in ecosystem states with long-lasting effects. The correct representation of these processes in ESMs allows them to reproduce the low-frequency IAV in vegetation, thereby extending the pf of GPP.

Although ESMs are similar in the fraction of GPP IAV that they can predict, there are substantial differences in the patterns and drivers of GPP IAV. The ESMs have distinct, non-overlapping hotspots of GPP IAV that drive the variability in atmospheric CO_2. We find large disparities in the role of Australia, southern Africa, and central South America in GPP IAV. The leading cause of the uncertainties in IAV patterns are differences in the response of GPP to soil moisture and the capability of the ESMs to simulate soil hydrology accurately. These differences materialize through the direct effect of soil moisture on photosynthesis and through the role of soil moisture on phenology. The inability of ESMs to reproduce GPP IAV also means that there are regions where the potential predictability of GPP does not resemble the actual predictive skill.

This study shows that the predictability of atmospheric CO_2 is currently not limited by the processes that provide predictability in the Earth system but rather by the representation of carbon flux variability patterns. The mismatches in GPP IAV imply that the IAV in atmospheric CO_2 is caused by different regions and by different drivers across the ESMs. Consequentially, when ESMs predict the atmospheric CO_2, the GPP anomalies that constitute the predicted CO_2 growth rate originate from different regions. Because the predicted CO_2 depends more on the distribution of GPP IAV hotspots than actual mechanisms that provide predictability, CO_2 forecast skill is not a suitable metric for studies on carbon flux predictability. An ESM with a high carbon flux predictability can be outperformed with respect to CO_2 forecast skill by a model that has a better representation of IAV patterns. A more suitable measure to assess carbon flux predictability could be the globally averaged anomaly correlation coefficient.

With the current uncertainties in GPP IAV patterns, the prediction of atmospheric CO_2 relies less on the prediction of regional climate anomalies and more on the predictable global climate patterns like ENSO. These global climate anomalies are able to balance out the regional differences in GPP IAV patterns by affecting large parts of the land surface simultaneously. In order to utilize the benefits of regional climate predictability for the predictability of CO_2, further work ought to focus on constraining GPP IAV and not on the processes providing predictability. The most limiting aspect in the use of ESMs to predict atmospheric CO_2 is a better understanding of the drivers of carbon flux variability. Whether GPP is limited by moisture, temperature, or radiation does not only affect variability patterns but also the predictability of the fluxes. An overestimation of humidity in an ecosystem by an ESM would result in GPP being more controlled by radiation than soil moisture, leading to an underestimation of predictability (or vice versa for systems that are too dry).

The findings of this study also suggest that previous estimations of ESM-based CO_2 forecast skill are underestimating the predictive capabilities of these systems. Various post-processing strategies could help to produce a CO_2 forecast skill that is not obscured by inaccurate IAV patterns but is a closer representation of the actual performance of ESM-based prediction systems. These strategies could include the scaling of carbon flux IAV patterns to resemble the observed IAV patterns. As there are strong regional differences in the predictive performance among the ESMs, another strategy would be to combine ESM predictions in a way that utilizes these differences. This could be done in a regionally weighted multi-model approach.

The limiting factor to predicting atmospheric CO_2 is the chaotic nature of weather and climate. However, our results show that we have not reached this limitation yet and that we are instead constrained by our understanding of terrestrial carbon flux variability. The development of observational products for terrestrial carbon fluxes, especially in the tropics, remains the main objective on the path of improving

the predictability of the global carbon cycle and atmospheric CO$_2$.

Code and data availability. The data and code used to produce the figures shown in this study are available at https://hdl.handle.net/21. 11116/0000-000D-72B4-7 (Dunkl et al., 2023). CE5

Author contributions. The study was conceptualized by ID, TI, and VB; ID developed the methodology, ran the analysis, and wrote the original draft; NL, AC, VKA, TI, and VB reviewed and edited the manuscript; and VB and TI provided supervision.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgements. The authors wish to thank the modelling groups that participated in the DCPP, SMYLE, and MiKlip projects. We are also grateful to Jürgen Bader for reviewing the manuscript.

Financial support. CE6 This project has received funding from the European Union's Horizon 2020 Research and Innovation programme (4C project, grant agreement no. 821003) and from the US National Science Foundation (grant no. OCE-1752724). hack

The article processing charges for this open-access publication were covered by the Max Planck Society.

Review statement. This paper was edited by David Medvigy and reviewed by two anonymous referees.

References

Ahlström, A., Raupach, M. R., Schurgers, G., Smith, B., Arneth, A., Jung, M., Reichstein, M., Canadell, J. G., Friedlingstein, P., Jain, A. K., Kato, E., Poulter, B., Sitch, S., Stocker, B. D., Viovy, N., Wang, Y. P., Wiltshire, A., Zaehle, S., and Zeng, N.: The dominant role of semi-arid ecosystems in the trend and variability of the land CO$_2$ sink, Science, 348, 895–899, https://doi.org/10.1126/science.aaa1668, 2015.

Alessandri, A. and Navarra, A.: On the coupling between vegetation and rainfall inter-annual anomalies: Possible contributions to seasonal rainfall predictability over land areas, Geophys. Res. Lett., 35, L02718, https://doi.org/10.1029/2007GL032415, 2008.

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., Murray-Tortarolo, G., Papale, D., Parazoo, N. C.,

Peylin, P., Piao, S., Sitch, S., Viovy, N., Wiltshire, A., and Zhao, M.: Spatiotemporal patterns of terrestrial gross primary production: A review, Rev. Geophys., 53, 785–818, https://doi.org/10.1002/2015RG000483, 2015.

Balmaseda, M. A., Mogensen, K., and Weaver, A. T.: Evaluation of the ECMWF ocean reanalysis system ORAS4, Q. J. Roy. Meteor. Soc., 139, 1132–1161, https://doi.org/10.1002/qj.2063, 2013.

Banzon, V., Smith, T. M., Chin, T. M., Liu, C., and Hankins, W.: A long-term record of blended satellite and in situ sea-surface temperature for climate monitoring, modeling and environmental studies, Earth Syst. Sci. Data, 8, 165–176, https://doi.org/10.5194/essd-8-165-2016, 2016.

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

Bellucci, A., Haarsma, R., Bellouin, N., Booth, B., Cagnazzo, C., van den Hurk, B., Keenlyside, N., Koenigk, T., Massonnet, F., Materia, S., and Weiss, M.: Advancements in decadal climate predictability: The role of nonoceanic drivers, Rev. Geophys., 53, 165–202, https://doi.org/10.1002/2014RG000473, 2015.

Betts, R. A., Jones, C. D., Knight, J. R., Keeling, R. F., and Kennedy, J. J.: El Niño and a record CO$_2$ rise, Nat. Clim. Change, 6, 806–810, https://doi.org/10.1038/nclimate3063, 2016.

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R., Mueller, W. A., Taylor, K. E., Zwiers, F., Rixen, M., Ruprich-Robert, Y., and Eade, R.: The Decadal Climate Prediction Project (DCPP) contribution to CMIP6, Geosci. Model Dev., 9, 3751–3777, https://doi.org/10.5194/gmd-9-3751-2016, 2016.

Bonan, G. B., Lombardozzi, D. L., Wieder, W. R., Oleson, K. W., Lawrence, D. M., Hoffman, F. M., and Collier, N.: Model Structure and Climate Data Uncertainty in Historical Simulations of the Terrestrial Carbon Cycle (1850–2014), Global Biogeochem. Cy., 33, 1310–1326, https://doi.org/10.1029/2019GB006175, 2019.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., Lavergne, C. d., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E.,Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Guez, E., L., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A. K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., and Vuichard, N.: Presentation and Evaluation of the IPSL-CM6A-LR Cli-

Proof only

mate Model, J. Adv. Model. Earth Sy., 12, e2019MS002010, https://doi.org/10.1029/2019MS002010, 2020.

Chen, M., Rafique, R., Asrar, G. R., Bond-Lamberty, B., Ciais, P., Zhao, F., Reyer, C. P. O., Ostberg, S., Chang, J., Ito, A., Yang, J., Zeng, N., Kalnay, E., West, T., Leng, G., Francois, L., Munhoven, G., Henrot, A., Tian, H., Pan, S., Nishina, K., Viovy, N., Morfopoulos, C., Betts, R., Schaphoff, S., Steinkamp, J., and Hickler, T.: Regional contribution to variability and trends of global gross primary productivity, Environ. Res. Lett., 12, 105005, https://doi.org/10.1088/1748-9326/aa8978, 2017.

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., Masina, S., Scoccimarro, E., Materia, S., Bellucci, A., and Navarra, A.: Global Mean Climate and Main Patterns of Variability in the CMCC-CM2 Coupled Model, J. Adv. Model. Earth Sy., 11, 185–209, https://doi.org/10.1029/2018MS001369, 2019.

Cheruy, F., Ducharne, A., Hourdin, F., Musat, I., Vignon, É., Gastineau, G., Bastrikov, V., Vuichard, N., Diallo, B., Dufresne, J.-L., Ghattas, J., Grandpeix, J.-Y., Idelkadi, A., Mellul, L., Maignan, F., Ménégoz, M., Ottlé, C., Peylin, P., Servonnat, J., Wang, F., and Zhao, Y.: Improved Near-Surface Continental Climate in IPSL-CM6A-LR by Combined Evolutions of Atmospheric and Land Surface Physics, J. Adv. Model. Earth Sy., 12, e2019MS002005, https://doi.org/10.1029/2019MS002005, 2020.

Chikamoto, Y., Timmermann, A., Widlansky, M. J., Balmaseda, M. A., and Stott, L.: Multi-year predictability of climate, drought, and wildfire in southwestern North America, Sci. Rep.-UK, 7, 6568, https://doi.org/10.1038/s41598-017-06869-7, 2017.

Collalti, A., Ibrom, A., Stockmarr, A., Cescatti, A., Alkama, R., Fernández-Martínez, M., Matteucci, G., Sitch, S., Friedlingstein, P., Ciais, P., Goll, D. S., Nabel, J. E. M. S., Pongratz, J., Arneth, A., Haverd, V., and Prentice, I. C.: Forest production efficiency increases with growth temperature, Nat. Commun., 11, 5322, https://doi.org/10.1038/s41467-020-19187-w, 2020.

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., Mu, M., and Randerson, J. T.: The International Land Model Benchmarking (ILAMB) System: Design, Theory, and Implementation, J. Adv. Model. Earth Sy., 10, 2731–2754, https://doi.org/10.1029/2018MS001354, 2018.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., Kampenhout, L. v., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model Version 2 (CESM2), J. Adv. Model. Earth Sy., 12, e2019MS001916, https://doi.org/10.1029/2019MS001916, 2020.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., Berg, L. v. d., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., Rosnay, P. d., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Dirkson, A., Merryfield, W. J., and Monahan, A.: Impacts of Sea Ice Thickness Initialization on Seasonal Arctic Sea Ice Predictions, J. Climate, 30, 1001–1017, https://doi.org/10.1175/JCLI-D-16-0437.1, 2017.

Dirmeyer, P. A., Halder, S., and Bombardi, R.: On the Harvest of Predictability From Land States in a Global Forecast Model, J. Geophys. Res.-Atmos., 123, 13111–13127, https://doi.org/10.1029/2018JD029103, 2018.

Dunkl, I., Spring, A., Friedlingstein, P., and Brovkin, V.: Process-based analysis of terrestrial carbon flux predictability, Earth Syst. Dynam., 12, 1413–1426, https://doi.org/10.5194/esd-12-1413-2021, 2021.

Dunkl, I., Lovenduski, N., Collalti, A., Arora, V. K., Ilyina, T., and Brovkin, V.: GPP and the predictability of CO$_2$: more uncertainty in what we predict than how well we predict it, MPG PuRe [code, data set], https://hdl.handle.net/21.11116/0000-000D-72B4-7, 2023. TS1

Estella-Perez, V., Mignot, J., Guilyardi, E., Swingedouw, D., and Reverdin, G.: Advances in reconstructing the AMOC using sea surface observations of salinity, Clim. Dynam., 55, 975–992, https://doi.org/10.1007/s00382-020-05304-4, 2020.

Feng, X., DelSole, T., and Houser, P.: Bootstrap estimated seasonal potential predictability of global temperature and precipitation, Geophys. Res. Lett., 38, L07702, https://doi.org/10.1029/2010GL046511, 2011.

Gavrilov, M. B., An, W., Xu, C., Radaković, M. G., Hao, Q., Yang, F., Guo, Z., Perić, Z., Gavrilov, G., and Marković, S. B.: Independent Aridity and Drought Pieces of Evidence Based on Meteorological Data and Tree Ring Data in Southeast Banat, Vojvodina, Serbia, Atmosphere, 10, 586, https://doi.org/10.3390/atmos10100586, 2019.

Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H.-D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., Mikolajewicz, U., Mueller, W., Notz, D., Pithan, F., Raddatz, T., Rast, S., Redler, R., Roeckner, E., Schmidt, H., Schnur, R., Segschneider, J., Six, K. D., Stockhause, M., Timmreck, C., Wegner, J., Widmann, H., Wieners, K.-H., Claussen, M., Marotzke, J., and Stevens, B.: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, J. Adv. Model. Earth Sy., 5, 572–597, https://doi.org/10.1002/jame.20038, 2013.

Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, J. Geophys. Res.-Oceans, 118, 6704–6716, https://doi.org/10.1002/2013JC009067, 2013.

Holmgren, M., Scheffer, M., Ezcurra, E., Gutiérrez, J. R., and Mohren, G. M. J.: El Niño effects on the dynamics of terrestrial ecosystems, Trends Ecol. Evol., 16, 89–94, https://doi.org/10.1016/S0169-5347(00)02052-8, 2001.

Hu, Q., Li, T., Deng, X., Wu, T., Zhai, P., Huang, D., Fan, X., Zhu, Y., Lin, Y., Xiao, X., Chen, X., Zhao, X., Wang, L., and Qin, Z.: Intercomparison of global terrestrial carbon fluxes estimated

by MODIS and Earth system models, Sci. Total Environ., 810, 152231, https://doi.org/10.1016/j.scitotenv.2021.152231, 2022.

Hurrell, J. W., Holland, M. M., Gent, P. R., Ghan, S., Kay, J. E., Kushner, P. J., Lamarque, J.-F., Large, W. G., Lawrence, D., Lindsay, K., Lipscomb, W. H., Long, M. C., Mahowald, N., Marsh, D. R., Neale, R. B., Rasch, P., Vavrus, S., Vertenstein, M., Bader, D., Collins, W. D., Hack, J. J., Kiehl, J., and Marshall, S.: The Community Earth System Model: A Framework for Collaborative Research, B. Am. Meteorol. Soc., 94, 1339–1360, https://doi.org/10.1175/BAMS-D-12-00121.1, 2013.

Ilyina, T., Li, H., Spring, A., Müller, W. A., Bopp, L., Chikamoto, M. O., Danabasoglu, G., Dobrynin, M., Dunne, J., Fransner, F., Friedlingstein, P., Lee, W., Lovenduski, N. S., Merryfield, W. J., Mignot, J., Park, J. Y., Séférian, R., Sospedra-Alfonso, R., Watanabe, M., and Yeager, S.: Predictable Variations of the Carbon Sinks and Atmospheric CO$_2$ Growth in a Multi-Model Framework, Geophys. Res. Lett., 48, e2020GL090695, https://doi.org/10.1029/2020GL090695, 2021.

Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., Gans, F., Ichii, K., Jain, A. K., Kato, E., Papale, D., Poulter, B., Raduly, B., Rödenbeck, C., Tramontana, G., Viovy, N., Wang, Y.-P., Weber, U., Zaehle, S., and Zeng, N.: Compensatory water effects link yearly global land CO$_2$ sink changes to temperature, Nature, 541, 516–520, https://doi.org/10.1038/nature20780, 2017.

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

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J. M., Bates, S. C., Danabasoglu, G., Edwards, J., Holland, M., Kushner, P., Lamarque, J.-F., Lawrence, D., Lindsay, K., Middleton, A., Munoz, E., Neale, R., Oleson, K., Polvani, L., and Vertenstein, M.: The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability, B. Am. Meteorol. Soc., 96, 1333–1349, https://doi.org/10.1175/BAMS-D-13-00255.1, 2015.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., and Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, J. Meteorol. Soc. Jpn. Ser. II, 93, 5–48, https://doi.org/10.2151/jmsj.2015-001, 2015.

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

Kumar, S., Newman, M., Lawrence, D. M., Lo, M.-H., Akula, S., Lan, C.-W., Livneh, B., and Lombardozzi, D.: The GLACE-Hydrology Experiment: Effects of Land–Atmosphere Coupling on Soil Moisture Variability and Predictability, J. Climate, 33, 6511–6529, https://doi.org/10.1175/JCLI-D-19-0598.1, 2020.

Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., Zeng, X., Yang, Z.-L., Levis, S., Sakaguchi, K., Bonan, G. B., and Slater, A. G.: Parameterization improvements and functional and structural advances in Version 4 of the Community Land Model, J. Adv. Model. Earth Sy., 3, M03001, https://doi.org/10.1029/2011MS00045, 2011.

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B., Kampenhout, L. v., Kennedy, D., Kluzek, E., Lawrence, P. J., Li, F., Li, H., Lombardozzi, D., Riley, W. J., Sacks, W. J., Shi, M., Vertenstein, M., Wieder, W. R., Xu, C., Ali, A. A., Badger, A. M., Bisht, G., Broeke, M. v. d., Brunke, M. A., Burns, S. P., Buzan, J., Clark, M., Craig, A., Dahlin, K., Drewniak, B., Fisher, J. B., Flanner, M., Fox, A. M., Gentine, P., Hoffman, F., Keppel-Aleks, G., Knox, R., Kumar, S., Lenaerts, J., Leung, L. R., Lipscomb, W. H., Lu, Y., Pandey, A., Pelletier, J. D., Perket, J., Randerson, J. T., Ricciuto, D. M., Sanderson, B. M., Slater, A., Subin, Z. M., Tang, J., Thomas, R. Q., Martin, M. V., and Zeng, X.: The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty, J. Adv. Model. Earth Sy., 11, 4245–4287, https://doi.org/10.1029/2018MS001583, 2019.

Li, X. and Xiao, J.: Mapping Photosynthesis Solely from Solar-Induced Chlorophyll Fluorescence: A Global, Fine-Resolution Dataset of Gross Primary Production Derived from OCO-2, Remote Sens.-Basel, 11, 2563, https://doi.org/10.3390/rs11212563, 2019.

Lovato, T., Peano, D., Butenschön, M., Materia, S., Iovino, D., Scoccimarro, E., Fogli, P. G., Cherchi, A., Bellucci, A., Gualdi, S., Masina, S., and Navarra, A.: CMIP6 Simulations With the CMCC Earth System Model (CMCC-ESM2), J. Adv. Model. Earth Sy., 14, e2021MS002814, https://doi.org/10.1029/2021MS002814, 2022.

Lovenduski, N. S., Bonan, G. B., Yeager, S. G., Lindsay, K., and Lombardozzi, D. L.: High predictability of terrestrial carbon fluxes from an initialized decadal prediction system, Environ. Res. Lett., 14, 124074, https://doi.org/10.1088/1748-9326/ab5c55, 2019.

Marotzke, J., Müller, W. A., Vamborg, F. S. E., Becker, P., Cubasch, U., Feldmann, H., Kaspar, F., Kottmeier, C., Marini, C., Polkova, I., Prömmel, K., Rust, H. W., Stammer, D., Ulbrich, U., Kadow, C., Köhl, A., Kröger, J., Kruschke, T., Pinto, J. G., Pohlmann, H., Reyers, M., Schröder, M., Sienz, F., Timmreck, C., and Ziese, M.: MiKlip: A National Research Project on Decadal Climate Prediction, B. Am. Meteorol. Soc., 97, 2379–2394, https://doi.org/10.1175/BAMS-D-15-00184.1, 2016.

Meehl, G. A., Richter, J. H., Teng, H., Capotondi, A., Cobb, K., Doblas-Reyes, F., Donat, M. G., England, M. H., Fyfe, J. C., Han, W., Kim, H., Kirtman, B. P., Kushnir, Y., Lovenduski, N. S., Mann, M. E., Merryfield, W. J., Nieves, V., Pegion, K., Rosenbloom, N., Sanchez, S. C., Scaife, A. A., Smith, D., Subramanian, A. C., Sun, L., Thompson, D., Ummenhofer, C. C., and Xie, S.-P.: Initialized Earth System prediction from subseasonal to decadal timescales, Nature Reviews Earth & Environment, 2, 340–357, https://doi.org/10.1038/s43017-021-00155-x, 2021.

Muggeo, V.: Segmented: An R Package to Fit Regression Models With Broken-Line Relationships, R News, 8, 20–25, 2008.

Nicolì, D., Bellucci, A., Ruggieri, P., Athanasiadis, P. J., Materia, S., Peano, D., Fedele, G., Hénin, R., and Gualdi, S.: The Euro-Mediterranean Center on Climate Change (CMCC) decadal prediction system, Geosci. Model Dev., 16, 179–197, https://doi.org/10.5194/gmd-16-179-2023, 2023.

O'Sullivan, M., Smith, W. K., Sitch, S., Friedlingstein, P., Arora, V. K., Haverd, V., Jain, A. K., Kato, E., Kautz, M., Lombardozzi, D., Nabel, J. E. M. S., Tian, H., Vuichard, N., Wiltshire, A., Zhu, D., and Buermann, W.: Climate-Driven Variability and Trends in Plant Productivity Over Recent Decades Based on Three Global Products, Global Biogeochem. Cy., 34, e2020GB006613, https://doi.org/10.1029/2020GB006613, 2020.

Padrón, R. S., Gudmundsson, L., Liu, L., Humphrey, V., and Seneviratne, S. I.: Drivers of intermodel uncertainty in land carbon sink projections, Biogeosciences, 19, 5435–5448, https://doi.org/10.5194/bg-19-5435-2022, 2022.

Peano, D., Materia, S., Collalti, A., Alessandri, A., Anav, A., Bombelli, A., and Gualdi, S.: Global Variability of Simulated and Observed Vegetation Growing Season, J. Geophys. Res.-Biogeo., 124, 3569–3587, https://doi.org/10.1029/2018JG004881, 2019.

Peano, D., Hemming, D., Materia, S., Delire, C., Fan, Y., Joetzjer, E., Lee, H., Nabel, J. E. M. S., Park, T., Peylin, P., Wårlind, D., Wiltshire, A., and Zaehle, S.: Plant phenology evaluation of CRESCENDO land surface models – Part 1: Start and end of the growing season, Biogeosciences, 18, 2405–2428, https://doi.org/10.5194/bg-18-2405-2021, 2021.

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

Qiao, L., Zuo, Z., and Xiao, D.: Evaluation of Soil Moisture in CMIP6 Simulations, J. Climate, 35, 779–800, https://doi.org/10.1175/JCLI-D-20-0827.1, 2022.

Reick, C. H., Raddatz, T., Brovkin, V., and Gayler, V.: Representation of natural and anthropogenic land cover change in MPI-ESM, J. Adv. Model. Earth Sy., 5, 459–482, https://doi.org/10.1002/jame.20022, 2013.

Santanello, J. A., Dirmeyer, P. A., Ferguson, C. R., Findell, K. L., Tawfik, A. B., Berg, A., Ek, M., Gentine, P., Guillod, B. P., Heerwaarden, C. v., Roundy, J., and Wulfmeyer, V.: Land–Atmosphere Interactions: The LoCo Perspective, B. Am. Meteorol. Soc., 99, 1253–1272, https://doi.org/10.1175/BAMS-D-17-0001.1, 2018.

Séférian, R., Berthet, S., and Chevallier, M.: Assessing the Decadal Predictability of Land and Ocean Carbon Uptake, Geophys. Res. Lett., 45, 2455–2466, https://doi.org/10.1002/2017GL076092, 2018.

Seo, E., Lee, M.-I., Jeong, J.-H., Koster, R. D., Schubert, S. D., Kim, H.-M., Kim, D., Kang, H.-S., Kim, H.-K., MacLachlan, C., and Scaife, A. A.: Impact of soil moisture initialization on boreal summer subseasonal forecasts: mid-latitude surface air temperature and heat wave events, Clim. Dynam., 52, 1695–1709, https://doi.org/10.1007/s00382-018-4221-4, 2019.

Smith, T. M., Reynolds, R. W., Peterson, T. C., and Lawrimore, J.: Improvements to NOAA's Historical Merged Land–Ocean Surface Temperature Analysis (1880–2006), J. Climate, 21, 2283–2296, https://doi.org/10.1175/2007JCLI2100.1, 2008.

Song, X., Wang, D.-Y., Li, F., and Zeng, X.-D.: Evaluating the performance of CMIP6 Earth system models in simulating global vegetation structure and distribution, Advances in Climate Change Research, 12, 584–595, https://doi.org/10.1016/j.accre.2021.06.008, 2021.

Spring, A. and Ilyina, T.: Predictability Horizons in the Global Carbon Cycle Inferred From a Perfect-Model Framework, Geophys. Res. Lett., 47, e2019GL085311, https://doi.org/10.1029/2019GL085311, 2020.

Storto, A. and Masina, S.: C-GLORSv5: an improved multipurpose global ocean eddy-permitting physical reanalysis, Earth Syst. Sci. Data, 8, 679–696, https://doi.org/10.5194/essd-8-679-2016, 2016.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Sigmond, M., Solheim, L., von Salzen, K., Yang, D., and Winter, B.: The Canadian Earth System Model version 5 (CanESM5.0.3), Geosci. Model Dev., 12, 4823–4873, https://doi.org/10.5194/gmd-12-4823-2019, 2019.

Titchner, H. A. and Rayner, N. A.: The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations, J. Geophys. Res.-Atmos., 119, 2864–2889, https://doi.org/10.1002/2013JD020316, 2014.

Tsujino, H., Urakawa, S., Nakano, H., Small, R. J., Kim, W. M., Yeager, S. G., Danabasoglu, G., Suzuki, T., Bamber, J. L., Bentsen, M., Böning, C. W., Bozec, A., Chassignet, E. P., Curchitser, E., Boeira Dias, F., Durack, P. J., Griffies, S. M., Harada, Y., Ilicak, M., Josey, S. A., Kobayashi, C., Kobayashi, S., Komuro, Y., Large, W. G., Le Sommer, J., Marsland, S. J., Masina, S., Scheinert, M., Tomita, H., Valdivieso, M., and Yamazaki, D.: JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do), Ocean Model., 130, 79–139, https://doi.org/10.1016/j.ocemod.2018.07.002, 2018.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. a. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Q. J. Roy. Meteor. Soc., 131, 2961–3012, https://doi.org/10.1256/qj.04.176, 2005.

Wang, G., Sun, S., and Mei, R.: Vegetation dynamics contributes to the multi-decadal variability of precipitation in the Amazon region, Geophys. Res. Lett., 38, L19703, https://doi.org/10.1029/2011GL049017, 2011.

Wang, J., Zeng, N., and Wang, M.: Interannual variability of the atmospheric CO$_2$ growth rate: roles of precipitation and temperature, Biogeosciences, 13, 2339–2352, https://doi.org/10.5194/bg-13-2339-2016, 2016.

Wigneron, J.-P., Fan, L., Ciais, P., Bastos, A., Brandt, M., Chave, J., Saatchi, S., Baccini, A., and Fensholt, R.: Tropical forests did not recover from the strong 2015–2016 El Niño event, Science Advances, 6, eaay4603, https://doi.org/10.1126/sciadv.aay4603, 2020.

Wu, R.-J., Lo, M.-H., and Scanlon, B. R.: The Annual Cycle of Terrestrial Water Storage Anomalies in CMIP6 Models Evaluated against GRACE Data, J. Climate, 34, 8205–8217, https://doi.org/10.1175/JCLI-D-21-0021.1, 2021.

Xue, Y., Smith, T. M., and Reynolds, R. W.: Interdecadal Changes of 30-Yr SST Normals during 1871–2000,

J. Climate, 16, 1601–1612, https://doi.org/10.1175/1520-0442(2003)016<1601:ICOYSN>2.0.CO;2, 2003.

Yang, C., Masina, S., and Storto, A.: Historical ocean reanalyses (1900–2010) using different data assimilation strategies, Q. J. Roy. Meteor. Soc., 143, 479–493, https://doi.org/10.1002/qj.2936, 2017.

Yeager, S. G., Danabasoglu, G., Rosenbloom, N. A., Strand, W., Bates, S. C., Meehl, G. A., Karspeck, A. R., Lindsay, K., Long, M. C., Teng, H., and Lovenduski, N. S.: Predicting Near-Term Changes in the Earth System: A Large Ensemble of Initialized Decadal Prediction Simulations Using the Community Earth System Model, B. Am. Meteorol. Soc., 99, 1867–1886, https://doi.org/10.1175/BAMS-D-17-0098.1, 2018.

Yeager, S. G., Rosenbloom, N., Glanville, A. A., Wu, X., Simpson, I., Li, H., Molina, M. J., Krumhardt, K., Mogen, S., Lindsay, K., Lombardozzi, D., Wieder, W., Kim, W. M., Richter, J. H., Long, M., Danabasoglu, G., Bailey, D., Holland, M., Lovenduski, N., Strand, W. G., and King, T.: The Seasonal-to-Multiyear Large Ensemble (SMYLE) prediction system using the Community Earth System Model version 2, Geosci. Model Dev., 15, 6451–6493, https://doi.org/10.5194/gmd-15-6451-2022, 2022.

Zeng, N., Yoon, J.-H., Vintzileos, A., Collatz, G. J., Kalnay, E., Mariotti, A., Kumar, A., Busalacchi, A., and Lord, S.: Dynamical prediction of terrestrial ecosystems and the global carbon cycle: A 25-year hindcast experiment, Global Biogeochem. Cy., 22, GB4015, https://doi.org/10.1029/2008GB003183, 2008.

Zhang, Y. and Ye, A.: Would the obtainable gross primary productivity (GPP) products stand up? A critical assessment of 45 global GPP products, Sci. Total Environ., 783, 146965, https://doi.org/10.1016/j.scitotenv.2021.146965, 2021.

Zhang, Y., Xiao, X., Wu, X., Zhou, S., Zhang, G., Qin, Y., and Dong, J.: A global moderate resolution dataset of gross primary production of vegetation for 2000–2016, Scientific Data, 4, 170165, https://doi.org/10.1038/sdata.2017.165, 2017.

Zhang, Y., Dannenberg, M. P., Hwang, T., and Song, C.: El Niño-Southern Oscillation-Induced Variability of Terrestrial Gross Primary Production During the Satellite Era, J. Geophys. Res.-Biogeo., 124, 2419–2431, https://doi.org/10.1029/2019JG005117, 2019.

Zhang, Y., Keenan, T. F., and Zhou, S.: Exacerbated drought impacts on global ecosystems due to structural overshoot, Nature Ecology & Evolution, 5, 1490–1498, https://doi.org/10.1038/s41559-021-01551-8, 2021.

Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., and Mayer, M.: The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: a description of the system and assessment, Ocean Sci., 15, 779–808, https://doi.org/10.5194/os-15-779-2019, 2019.

Remarks from the language copy-editor

CE1   Does this affiliation refer to two separate institutes? If so, according to our house standards, it must be split into two separate affiliations. Please advise.

CE2   Please confirm the change based on your feedback during proofreading.

CE3   Please confirm the change.

CE4   Please note that my question here was regarding the placement of "10^{-13}" (although I am also very glad that the scaling error came to light). My question was if the unit should perhaps be "$\mathrm{kg\,C}\,10^{-9}\,\mathrm{m}^{-2}\,\mathrm{s}^{-1}$"? Please note that changing the scaling factor and replacing Figs. 4 and 6 is not a problem, but it will require approval from the editor. Please provide a detailed explanation of why these changes are required, and we will ask the handling editor for approval. Thank you.

CE5   Please note the slight edit.

CE6   Please note the slight edits.

Remarks from the typesetter

TS1   Please confirm.