



# Global Evaluation of Gross Primary Productivity in the JULES Land Surface Model.

Darren Slevin<sup>1</sup>, Simon F. B. Tett<sup>1</sup>, Jean-François Exbrayat<sup>1, 2</sup>, A. Anthony Bloom<sup>1, 2, 3</sup>, and Mathew Williams<sup>1, 2</sup>

<sup>1</sup>School of GeoSciences, The University of Edinburgh, Crew Building, Alexander Crum Brown Road, Edinburgh, EH9 3FF, UK

<sup>2</sup>National Centre for Earth Observation, The University of Edinburgh, Crew Building, Alexander Crum Brown Road, Edinburgh, EH9 3FF, UK

<sup>3</sup>Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

*Correspondence to:* Darren Slevin (d.slevin@ed.ac.uk)

**Abstract.** This study evaluates the ability of the JULES Land Surface Model (LSM) to simulate Gross Primary Productivity (GPP) at regional and global scales for 2001–2010. Model simulations, performed at various spatial resolutions and driven with a variety of meteorological datasets (WFDEI-GPCC, WFDEI-CRU and PRINCETON), were compared to the MODIS GPP product, spatially gridded estimates of upscaled GPP from the FLUXNET network (FLUXNET-MTE) and the CARDAMOM terrestrial carbon cycle analysis. Firstly, JULES was found to simulate interannual variability (IAV) at global scales. When JULES was driven with the WFDEI-GPCC dataset (at  $0.5^\circ \times 0.5^\circ$  spatial resolution), it was found that the annual average global GPP simulated by JULES for 2001–2010 was higher than the observation-based estimates (MODIS and FLUXNET-MTE), by 25 % and 8 %, respectively, and CARDAMOM estimates by 23 %. Secondly, GPP fluxes simulated by JULES for various biomes (forests, grasslands and shrubs) at global and regional scales were compared. It was found that differences between JULES, MODIS, FLUXNET-MTE and CARDAMOM at global scales were mostly due to differences in the tropics with CARDAMOM performing better than JULES in this region. Thirdly, it was shown that spatial resolution ( $0.5^\circ \times 0.5^\circ$ ,  $1^\circ \times 1^\circ$  and  $2^\circ \times 2^\circ$ ) had no impact on simulated GPP on these large scales. Finally, the sensitivity of JULES to meteorological driving data, a major source of model uncertainty, was examined. Estimates of annual average global GPP were higher when JULES was driven with the PRINCETON meteorological dataset than when driven with the WFDEI-GPCC dataset by  $4 \text{ Pg C year}^{-1}$ . At regional scales, differences between two were observed with the WFDEI-GPCC driven model simulations estimating higher GPP in the tropics (at  $5^\circ\text{N}–5^\circ\text{S}$ ) and the PRINCETON driven model simulations estimating higher GPP in the extratropics (at  $30^\circ\text{N}–60^\circ\text{N}$ ).

## 1 Introduction

Changes in atmospheric  $\text{CO}_2$  concentrations and water vapour can alter the energy balance of the atmosphere and thus climate. One important influence on these greenhouse gases is the land surface. The land surface is an important component of the climate system, provides the lower boundary for the atmosphere and exchanges energy, water and carbon (C) with the atmosphere



(Pielke et al., 1998; Pitman, 2003; Seneviratne and Stöckli, 2008). It also controls the partitioning of available energy (into latent and sensible heat) and water (into evaporation and runoff) at the surface, is the location of the terrestrial C sink and influences weather and climate and vice versa (Bonan, 2008). Changes in the land surface can influence climate at various time and spatial scales and since the land surface is the location of the terrestrial C cycle, it's ability to act as a C source or sink can influence atmospheric CO<sub>2</sub> concentrations (Le Quéré et al., 2009; Pan et al., 2011; Le Quéré et al., 2013; Tian et al., 2016).

The reduced ability of the land surface to absorb increased anthropogenic CO<sub>2</sub> emissions in the future has been shown by models and observations (Friedlingstein et al., 2006; Canadell et al., 2007; Friedlingstein et al., 2014; Sitch et al., 2015). The Coupled Climate–Carbon Cycle Model Intercomparison Project (Friedlingstein et al., 2006, C4MIP) and phase 5 of the Coupled Model Intercomparison Project (Arora et al., 2013; Friedlingstein et al., 2014, CMIP5) both showed a large spread in the future projections of atmospheric CO<sub>2</sub> by coupled climate carbon-cycle models using the same emission scenario. Using observations of atmospheric CO<sub>2</sub> concentrations since the 1960s, Canadell et al. (2007) showed a reduction in the efficiency of CO<sub>2</sub> sinks on land and oceans to store anthropogenic CO<sub>2</sub> emissions. Friedlingstein et al. (2006) and Friedlingstein et al. (2014) have suggested that a major source of model uncertainty is the land C cycle and this can affect the ability of earth system models (ESMs; also known as coupled carbon-cycle–climate models) to reliably simulate future atmospheric CO<sub>2</sub> concentrations and climate (Dalmonech et al., 2014).

Plants fix CO<sub>2</sub> as organic compounds through photosynthesis at the leaf scale and Gross Primary Productivity (GPP) is the total amount of C used in photosynthesis by plants at the ecosystem level (Beer et al., 2010; Chapin III et al., 2012). Photosynthesis at the leaf and canopy scale vary in response to changes in climate (temperature, precipitation, humidity and downward radiation fluxes) and nutrient availability (Anav et al., 2015). Terrestrial GPP is an important (and the largest) C flux since it drives several ecosystem functions such as respiration and growth (Beer et al., 2010). GPP contributes to the production of food, fibre, and wood for humans and along with respiration, is one of the major processes controlling the exchange of CO<sub>2</sub> between the land and atmosphere (Beer et al., 2010). It also plays an important role in the global C cycle helping terrestrial ecosystems to partially offset anthropogenic CO<sub>2</sub> emissions (Janssens et al., 2003; Cox and Jones, 2008; Battin et al., 2009; Anav et al., 2015)

However, at the global scale, there are no direct measurements of GPP (Anav et al., 2015). Global estimates of GPP exist, but are not solely based on measurements and, therefore, large uncertainties exist in these estimates (Anav et al., 2015). In LSMs, the correct simulation of GPP is important since errors in its calculation can propagate through the model and affect biomass and other flux calculations, such as Net Ecosystem Exchange (Schaefer et al., 2012, NEE). In JULES, NEE is not a model output and is calculated as total ecosystem respiration minus GPP. The correct representation of leaf level stomatal conductance influences GPP and transpiration and errors in GPP can also introduce errors into simulated latent and sensible heat fluxes.

Land surface models (LSMs) have become considerably more complex since the simple “bucket” model of Manabe (1969). Deardorff (1978) developed a model which could simulate temperature and moisture for two soil layers and included a vegetation layer. Sellers et al. (1986) built on the work of Deardorff (1978) by developing a globally applicable LSM. Foley et al. (1996) incorporated vegetation dynamics into an LSM. These developments have led to LSMs which can realistically represent



complex vegetation responses to meteorology, the climate effect of snow and biogeochemical processes (Pitman, 2003; van den Hurk et al., 2011). Therefore, as LSMs become more complex, their accuracy must be evaluated.

JULES has been evaluated at various scales: point (Blyth et al., 2010, 2011; Slevin et al., 2015; Ménard et al., 2015), regional (Galbraith et al., 2010; Burke et al., 2013; Chadburn et al., 2015) and globally as part of model-intercomparison studies (Anav et al., 2015; Sitch et al., 2015). Since biased simulations of GPP can introduce errors into other ecosystem processes in JULES, examining how the model performs at global and regional scales compared to GPP estimates from two observation-based datasets (FLUXNET-MTE and MODIS) and the CARbon DATA MOdel fraMework (Bloom et al., 2016, CARDAMOM) and how sources of uncertainty, such as the sensitivity of the model to spatial resolution and the meteorological data used to drive the model, affect model performance is an important part of model development.

10 In this study, the ability of JULES version 3.4.1 to simulate global and regional fluxes of GPP for various biomes, spatial resolutions and using different meteorological data to drive the model is evaluated. In particular, the following research questions are addressed:

- Can JULES capture interannual variability of GPP at the global scale? How do estimates of global GPP compare to those from observational-based datasets and the CARDAMOM framework?
- 15 – How does JULES GPP compare for various biomes at the global and regional scales?
- How sensitive are fluxes of GPP to the spatial resolution of the model?
- Is the meteorological data set used to drive the model important at the global scale?

## 2 Methods and model

### 2.1 Model description

20 The Joint UK Land Environment Simulator (JULES) is the land surface scheme of the UK Met Office Unified Model (MetUM, current version 10.2), which is a single model family used to simulate weather and climate across a range of timescales (Walters et al., 2014). JULES is a community land surface model which has evolved from the Met Office Surface Exchange Scheme (MOSES) (Cox et al., 1999) and is used for modelling all of the processes at the land surface, in the sub-surface soil, and surface exchange processes (Best et al., 2011; Clark et al., 2011). JULES can be used *offline* (i.e. outside of the host ESM, MetUM) and  
25 model simulations can be performed at point, regional and global scales. In LSMs, Plant Functional Types (PFTs) are used to represent broad groupings of plant species with similar ecosystem functions and resource use. In the version of JULES used in this study (version 3.4.1), each model gridbox consists of 9 different surface types; 5 PFTs (broadleaf trees, needleleaf trees, C3 (temperate) grass, C4 (tropical) grass and shrubs) and 4 non-vegetation surface types (urban, inland water, bare soil and land-ice). Model gridboxes can consist entirely of a mixture of the first 8 surface types or only land-ice. Since model version 4.2,  
30 each JULES gridbox can contain nine PFTS (tropical broadleaf evergreen, temperate broadleaf evergreen, broadleaf deciduous, needleleaf evergreen, needleleaf deciduous, C3, C4, evergreen shrub, deciduous shrub) (Harper et al., 2016).



The potential (without water and ozone stress) leaf-level photosynthesis ( $A_p$ ) is calculated using the  $C_3$  and  $C_4$  photosynthesis models of Collatz et al. (1991) and Collatz et al. (1992), respectively. This is calculated as the smoothed minimum of three limiting rates: (1) Rubisco-limited rate, (2) Light-limited rate and (3) Rate of transport of photosynthetic products ( $C_3$  plants) and PEP-Carboxylase limitation ( $C_4$  plants) (Clark et al., 2011). Leaf photosynthesis is linked to stomatal conductance via the  $CO_2$  diffusion equation using the Jacobs (1994) formulation. By taking soil moisture stress into account, leaf-level photosynthesis ( $A_l$ ) is calculated by multiplying the potential leaf-level photosynthesis ( $A_p$ ) by a soil moisture factor,  $\beta$  ( $A_l = A_p\beta$ ). The effect of  $O_3$  on leaf photosynthesis can also be included when calculating  $A_l$ , but it is not shown here.

There are two options available in JULES for radiation interception and the scaling of photosynthesis from leaf-level to canopy-level: (i) big leaf approach and (ii) multi-layer approach. For all model simulations performed in this study, the multi-layer approach was used. With the multi-layer approach, there are four variations (Table 3 of Clark et al. (2011)) that consider the vertical profile of canopy photosynthetic capacity, light inhibition of leaf respiration, the inclusion of sunfleck penetration and the division of canopy layers into sunlit and shaded leaves. Option 4 (from Table 3 of Clark et al. (2011)) was used for the model simulations performed in this study. This option includes the decrease in photosynthetic capacity from top to bottom of the canopy and the inhibition of leaf respiration in light. In the multi-layer approach, the amount of radiation absorbed and photosynthesis are estimated for a number of user defined canopy layers ( $dL_c = L_c/n$ , where  $L_c$  is the canopy leaf area and  $dL_c$  is the canopy layer leaf area). Canopy-scale fluxes are estimated to be the sum of the leaf-level fluxes in each canopy layer, scaled by leaf area. Clark et al. (2011) contains a more detailed description of leaf-level photosynthesis and its scaling up to the canopy-level.

Phenology (bud burst and leaf senescence) in JULES is usually updated once per day by multiplying the annual maximum LAI by a scaling factor (calculated using accumulated temperature-dependent leaf turnover rates). For each PFT, the C fluxes are calculated using a coupled photosynthesis-stomatal conductance model on each model timestep (typically 30 to 60 minutes) (Cox et al., 1998). These fluxes are then time-averaged (usually every 10 days) before being passed to TRIFFID (Top-down Representation of Interactive Foliage and Flora Including Dynamics), JULES' dynamic global vegetation model, which updates the vegetation distribution, based on the net C available to it and competition with other vegetation types, and soil C in each model gridbox on a longer timestep (usually every 10 days) (Cox, 2001). Biophysical properties of the land surface, such as albedo and roughness length, are updated based on the new vegetation state, so that these properties can affect the atmosphere (Clark et al., 2011). Clark et al. (2011) and Best et al. (2011) contain a more detailed description of JULES.

## 2.2 Experimental design

Offline simulations of GPP were performed at the global scale for the 2001–2010 period using various meteorological datasets and spatial resolutions (Table 1). The land cover was kept constant at values for the year 2000 (Loveland et al., 2000) and annual atmospheric  $CO_2$  concentrations were varied as in the historical record. In LSMs, the correct simulation of GPP is important since errors in its calculation can lead to errors in biomass and other land-atmosphere flux calculations, such as Net Ecosystem Exchange (NEE), and latent and sensible heat fluxes (Schaefer et al., 2012; Slevin et al., 2015). The 2001–2010 time period was used due to the availability of two global meteorological datasets used to drive JULES and estimates of



GPP, from two observation-based datasets and the CARDAMOM framework, used to evaluate model performance. JULES is being benchmarked against the upscaled FLUXNET, MODIS and CARDAMOM GPP.

Prior to performing the global scale model simulations, the soil moisture was brought to equilibrium using a 40 year global spin-up by cycling 10 years of meteorological data (1979–1989) twice and 10 years of meteorological data (1989–1999) twice (in equilibrium mode), followed by a 12 year spin-up by cycling 12 years of meteorological data (1999–2010) once (in dynamical mode). Finally, the actual model simulations were performed for 2001–2010 due to the availability of multiple global meteorological datasets and the observation-based (upscaled FLUXNET and MODIS) and CARDAMOM estimates of global GPP for this period. Clark et al. (2011) contains more information on spinning up the soil C pools.

### 2.3 Data

The datasets used in this study include those used as input to JULES (soil, vegetation and meteorological data) and the benchmarking data (FLUXNET-MTE, MODIS and CARDAMOM GPP) against which model performance is compared. The soil dataset used was the Harmonized World Soil Database version 1.2 (Nachtergaele et al., 2012, HWSD) and the land cover classification scheme used for specifying the PFT fractions for each model gridbox at the global scale was Global Land Cover Characterization database version 2.0 (Loveland et al., 2000, <http://edc2.usgs.gov/glcc/glcc.php>). The meteorological data used to drive JULES offline (i.e. run separately from its host Earth System Model) includes the downward shortwave and longwave radiation fluxes ( $\text{W m}^{-2}$ ), rainfall and snowfall rates ( $\text{kg m}^{-2} \text{s}^{-1}$ ), surface (2 m) air temperature (K), wind speed ( $\text{m s}^{-1}$ ), surface air pressure (Pa) and specific humidity ( $\text{kg kg}^{-1}$ ). Since JULES requires meteorological data at 6 hourly intervals or less in order to drive the model offline, a number of datasets were used; WFDEI (Weedon et al., 2014) and PRINCETON (Sheffield et al., 2006).

Global gridded estimates of GPP derived from the upscaling of observations from the FLUXNET network of tower sites (Jung et al., 2009), estimates from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor, aboard the U.S. National Aeronautics and Space Administration (NASA) Earth Observation System (EOS) satellites, Terra and Aqua (Yang et al., 2006), and GPP simulated by the CARbon DATA MOdel fraMework (Bloom et al., 2016, CARDAMOM) are used to evaluate model performance. These global gridded estimates of GPP provide a means to evaluate LSMs at large scales (Jung et al., 2009, 2010; Beer et al., 2010; Zhao and Running, 2010; Bonan et al., 2011; Lei et al., 2014).

#### 2.3.1 Forcing data

The WFDEI meteorological driving data was created using the same methodology as the WATCH Forcing Data (WFD) applied to the ERA-Interim reanalysis data for the 1979–2012 period (Weedon et al., 2014). The WFD data set (1901–2001) was created as part of the EU Water and Global Change (WATCH) project (Harding et al., 2011, <http://www.eu-watch.org/>) and was derived from the European Centre for Medium-range Weather Forecasts (ECMWF) ERA-40 reanalysis for 1958–2001 with data for 1901–1957 derived using randomly selected years from the ERA-40 data (Weedon et al., 2010, 2011). As part of the EMBRACE EU FP7 programme (<http://www.embrace-project.eu/>), the WFD methodology was applied to the ERA-Interim reanalysis data for the 1979–2013 period to generate the WFDEI meteorological forcing data (Weedon et al., 2014). As for the



WFD, WFDEI has two precipitation products, corrected using either CRU (Climate Research Unit at the University of East Anglia) or GPCC (Global Precipitation Climatology Centre) precipitation totals (Weedon et al., 2014) and are referred to as WFDEI-CRU and WFDEI-GPCC, respectively. The WFDEI dataset consists of 3 hourly, regularly gridded data at half-degree ( $0.5^\circ \times 0.5^\circ$ ) spatial resolution and is only available for land points including Antarctica. The dataset contains the following meteorological variables: downward shortwave and longwave radiation fluxes ( $\text{W m}^{-2}$ ), rainfall rate ( $\text{kg m}^{-2} \text{s}^{-1}$ ), snowfall rate ( $\text{kg m}^{-2} \text{s}^{-1}$ ), 2 m temperature (K), 10 m wind speed ( $\text{m s}^{-1}$ ), surface pressure (Pa) and 2 m specific humidity ( $\text{kg kg}^{-1}$ ).

The PRINCETON dataset is a global 62 year near-surface meteorological data set used for driving land surface models and was created by Princeton University's Terrestrial Hydrology Group (Sheffield et al., 2006, <http://hydrology.princeton.edu/home.php>). The PRINCETON data set consists of 3 hourly, regularly gridded data at 1-degree ( $1^\circ \times 1^\circ$ ) spatial resolution for the 1948–2010 period and is only available for land points excluding Antarctica. The dataset contains the following meteorological variables: downward shortwave and longwave radiation fluxes ( $\text{W m}^{-2}$ ), precipitation ( $\text{kg m}^{-2} \text{s}^{-1}$ ), air temperature (K), 10 m wind speed ( $\text{m s}^{-1}$ ), surface pressure (Pa) and specific humidity ( $\text{kg kg}^{-1}$ ).

### 2.3.2 Benchmarking data

The upscaled FLUXNET GPP (hereafter referred to as FLUXNET-MTE) was derived using a model tree ensemble (MTE) approach, a type of machine learning technique that can be trained to predict land-atmosphere fluxes (Jung et al., 2009) and provides a means to evaluate LSMs at large scales (Jung et al., 2009, 2010; Beer et al., 2010; Jung et al., 2011; Bonan et al., 2011). Based on observed meteorological data, land cover data and remotely sensed vegetation properties (fraction of absorbed photosynthetic active radiation), the upscaling principle uses a machine learning algorithm (model tree ensembles) to predict estimates of C fluxes at FLUXNET sites with available quality-filtered flux data and the trained model is then applied spatially using grids of the input data (Jung et al., 2009, 2011). However, these machine learning algorithms are typically data limited due to the quantity, quality and representativeness of the training dataset (Jung et al., 2009). There are two upscaled FLUXNET GPP datasets available depending on the flux partitioning method used to separate net ecosystem exchange of  $\text{CO}_2$  (NEE) into GPP and terrestrial ecosystem respiration (TER) (Reichstein et al., 2005; Lasslop et al., 2010). In this study, GPP based on the work by Reichstein et al. (2005) was used (this is the flux partitioning method used by the FLUXNET network). However, differences between the two upscaled FLUXNET GPP datasets are small. FLUXNET-MTE is a  $0.5^\circ \times 0.5^\circ$  spatial and monthly temporal resolution data set for the 1982–2011 period and is available from the Max Planck Institute for Biogeochemistry Data Portal (<https://www.bgc-jena.mpg.de/geodb/projects/Home.php>).

The MOD17 MODIS Gross/Net Primary Productivity (GPP/NPP) product provides continuous estimates of GPP/NPP for the Earth's entire land surface and is produced as part of the NASA's Earth Observing System (EOS) program. The MOD17 algorithm produces two subproducts, MOD17A2 (which stores 8-day composite GPP, net photosynthesis and QC flags) and MOD17A3 (annual NPP and QC flags) (Zhao et al., 2005). The resulting datasets contain regular gridded global estimates of GPP and NPP for the terrestrial land surface at the 1 km spatial resolution (Running et al., 2000). Sources of error in the Collection 4 primary production include mismatching spatial resolution between the gridded meteorological data ( $1^\circ \times 1.25^\circ$ ) and MODIS pixels (1 km), contaminated or missing 8-day LAI/FPAR (MOD15A2 MODIS product) due to cloud



cover or sensor malfunction and misclassified land cover from the MODIS land cover product (MOD12Q1) which can result in incorrect parameters from the MOD17 Biome Property Look-Up Table (BPLUT) and therefore leads to incorrect GPP estimates (Zhao et al., 2005). The Numerical Terradynamic Simulation Group (NTSG) (<http://www.ntsg.umt.edu/project/mod17>) at the University of Montana rectified these problems by spatial interpolation of the coarse resolution meteorological data, temporal  
5 infilling of cloud-contaminated MOD15A2 data and modification of BPLUT parameters based on observed GPP from flux tower measurements in order to create an improved MOD17 GPP product (Zhao et al., 2005). The global monthly MODIS GPP (version 55) dataset at  $0.05^\circ \times 0.05^\circ$  spatial resolution for the 2001–2010 period was downloaded from the NTSG ftp server ([ftp://ftp.ntsg.umt.edu/pub/MODIS/NTSG\\_Products/](ftp://ftp.ntsg.umt.edu/pub/MODIS/NTSG_Products/)). For the purposes of this study, the data was regridded to  $0.5^\circ \times 0.5^\circ$  spatial resolution.

10 The CARbon DAte MOdel fraMework (CARDAMOM) is a model-data fusion approach which consists of merging observational data with models in order to model quality and characterise its uncertainty (Bloom and Williams, 2015; Bloom et al., 2016). CARDAMOM relies on a Bayesian Markov Chain Monte Carlo (MCMC) algorithm to explore the parametric uncertainty of the ecosystem C balance model Data Assimilation Linked Ecosystem Carbon Model version two (Bloom et al., 2016, DALEC2) according to available C relevant data-streams (fluxes, leaf area index, changes in biomass, etc.). CARDAMOM can  
15 be applied at the point-scale and spatially with available remote-sensing based products such as MODIS LAI, biomass and soil carbon maps. When the framework is applied spatially, the Bayesian model-data fusion approach is performed in every model gridbox independently without using pre-defined biome maps. C fluxes, pool increments and parameter values with explicit confidence intervals attached to them are output from the MCMC algorithm. In this study, MODIS LAI, a tropical biomass map (Saatchi et al., 2011), a soil C dataset (Hiederer and Köchy, 2011), MODIS burned area (Giglio et al., 2013) and the  
20 ERA-Interim reanalysis data have been used as input to CARDAMOM in order to produce a global monthly mean GPP dataset at  $1^\circ \times 1^\circ$  spatial resolution for the 2001–2010 period (Bloom et al., 2016).

#### 2.4 Outline of experiments

This section describes the model simulations performed in this study (Table 1). For the JULES model simulations, the first part of the model simulation name refers to JULES version 3.4.1 and the second part refers to the global gridded meteorological  
25 dataset used to drive the model (Table 1). The spatial resolution of the model grid is appended to the end of the model simulation name. Model simulation names without an attached spatial resolution mean that the model simulation was performed at  $0.5^\circ \times 0.5^\circ$  spatial resolution. The TRIFFID dynamic global vegetation model (a submodel of JULES which simulates vegetation competition) and vegetation competition have been switched off for all JULES' simulations. Vegetation competition was disabled in order to prevent unrealistic vegetation fractions in model gridboxes for global scale simulations of GPP. For  
30 the CARDAMOM simulation, the ERA-Interim reanalysis product was used to drive the DALEC2 model at  $1^\circ \times 1^\circ$  resolution.



#### 2.4.1 Interannual variability of GPP

The ability of JULES to simulate the interannual variability (IAV) of GPP at the global scale was examined. Model simulations were performed for 2001–2010 using global parameter and meteorological datasets (JULES-WFDEI-GPCC; Table 1) with the results compared to GPP from observation-based estimates (FLUXNET-MTE and MODIS) and CARDAMOM.

#### 5 2.4.2 Global GPP

Model estimates of total annual GPP (JULES-WFDEI-GPCC) integrated across the globe were compared to FLUXNET-MTE, MODIS and CARDAMOM GPP. FLUXNET-MTE (global flux datasets derived from individual flux tower sites) and MODIS (satellite datasets) provide a means to evaluate JULES (and other LSMs) at global and regional scales (Bonan et al., 2011).

#### 2.4.3 Global and regional comparison for various biomes

- 10 In addition to deriving estimates of globally integrated GPP fluxes, the modelled (JULES-WFDEI-GPCC), FLUXNET-MTE, MODIS and CARDAMOM GPP were compared by biome type (Forest, Grassland and Shrub) at the global and regional scales (Global, Tropics and Extratropics). The global GPP was further analysed by biome type at the regional scale by dividing the global land area into seven regions (Figure 1; Table 2).

#### 2.4.4 Sensitivity to the spatial resolution of the input data

- 15 The sensitivity of the model to the spatial resolution of the input data was evaluated by varying the resolution of the ancillary data (soil and vegetation) and meteorological data (WFDEI-GPCC) and re-running the model simulations for 2001–2010. The input data was regridded from  $0.5^\circ \times 0.5^\circ$  to  $1^\circ \times 1^\circ$  spatial resolution (JULES-WFDEI-GPCC-1degree; Table 1) and from  $0.5^\circ \times 0.5^\circ$  to  $2^\circ \times 2^\circ$  spatial resolution (JULES-WFDEI-GPCC-2degree) using the first order conservative remapping function (remapcon) of the Climate Data Operators (CDO) software package (<https://code.zmaw.de/projects/cdo>). The observation-  
20 based (FLUXNET-MTE and MODIS) datasets were regridded using this method. The output from these simulations were compared to those at  $0.5^\circ \times 0.5^\circ$  spatial resolution (JULES-WFDEI-GPCC; Table 1).

#### 2.4.5 Sensitivity to the meteorological driving data set

- The sensitivity of JULES to the meteorological driving data was evaluated by comparing model simulations driven using the WFDEI-GPCC (JULES-WFDEI-GPCC-1degree; Table 1) and PRINCETON datasets (JULES-PRINCETON; Table 1) at  
25  $1^\circ \times 1^\circ$  spatial resolution. In these model simulations, the same ancillary datasets are used by both with the only difference in the model simulations being the meteorological data used to drive the model.



## 2.5 Model analyses

In order to quantify how the model performs at the global scale, the following metrics were used: global area-weighted mean ( $\bar{x}$ ; Equation 1), Coefficient of Variation (CV; Equation 2) and monthly anomalies (Equation 3).

$$\bar{x} = \frac{\sum_{i,j=1}^{i=m, j=n} a_{i,j} x_{i,j}}{\sum_{i,j=1}^{i=m, j=n} a_{i,j}} \quad (1)$$

5 The global area-weighted mean is calculated by multiplying the monthly GPP flux for each grid box ( $x_{i,j}$ ) by the area of its grid box ( $a_{i,j}$ ) and dividing the sum of these values by the total land surface area.  $m$  and  $n$  are the total number of grid boxes in the  $x$ - and  $y$ -direction, respectively. For example, when running a global scale model simulation at half-degree ( $0.5^\circ \times 0.5^\circ$ ) spatial resolution,  $m = 720$  (number of grid boxes in the west-east direction) and  $n = 360$  (number of grid boxes in the north-south direction).

$$10 \quad CV = \frac{\sigma}{\mu} \times 100 \quad (2)$$

CV (also known as relative variability) is a measure of the relative magnitude of the standard deviation ( $\sigma$ ) and is calculated by dividing the standard deviation by the mean ( $\mu$ ). It is expressed as a percentage and is always positive. CV is a useful statistic since it allows the degree of variation of various datasets to be compared even if the means are quite different from each other. It is also dimensionless which means that CVs can be used to compare the dispersion (variability) of the data when  
15 other measures like standard deviation or root mean squared error cannot.

To quantify model performance at the global scale, CV was calculated by first computing the standard deviation and means of the global area-weighted means for each month and then dividing the average of the standard deviations by the average of the means for each month.

$$\text{Monthly anomaly} = x - \bar{x}_{clim} \quad (3)$$

20 The monthly anomaly is defined as the departure of the observed monthly values ( $x$ ) from the long-term (climatological) average for that month ( $\bar{x}_{clim}$ ).

## 3 Results

### 3.1 Interannual variability of GPP

JULES simulates the seasonal cycle of GPP (JULES-WFDEI-GPCC; Table 1) at the global scale (Figure 2a) with the global  
25 area-weighted average of its monthly GPP for 2001–2010 falling within range of the observation-based estimates (FLUXNET-MTE and MODIS) for much of the year (between  $64$  and  $107 \text{ g C m}^{-2} \text{ month}^{-1}$ ). A similar trend can be found with the CARDAMOM GPP (Figure 2a). The exception to this are the winter months (January, February, March and December) with



JULES simulating higher global mean GPP by  $2 \text{ g C m}^{-2} \text{ month}^{-1}$  on average compared to FLUXNET-MTE. The MODIS GPP means are lower than FLUXNET-MTE for each of the monthly climatologies by  $10 \text{ g C m}^{-2} \text{ month}^{-1}$  on average (Figure 2a).

The standard deviation of the monthly GPP fluxes is used to measure interannual variability and this is expressed as a percentage of the mean monthly GPP fluxes using coefficient of variation (CV). The CV of the model simulated and observation-based GPP fluxes range between 0.8–4 % for the mean monthly GPP with the highest differences between the monthly values being for winter and spring (February, March, April, November and December) (Figure 2b). This pattern is similar to the global area-weighted average of the monthly climatologies (Figure 2a). Low values of CV mean that differences between the monthly GPP fluxes and the mean monthly GPP fluxes are small and larger CV values mean the opposite.

The monthly anomalies (computed using the global area-weighted mean values) expressed as percentages of the global area-weighted mean of the model simulated mean monthly GPP (JULES-WFDEI-GPCC) compare equally well to both FLUXNET-MTE and MODIS GPP for 2001–2010 with both having Root Mean Squared Errors (RMSEs) of 2.4 % with CARDAMOM having much lower year to year variation (Figure 2c). However, the high variation in JULES GPP at the beginning and end of the year is observed in the monthly anomalies from 2001 to 2010 which in some years, such as 2002, the model is unable to capture monthly GPP (Figure 2c).

Overall, it was found that JULES can simulate interannual variability at global scales.

### 3.2 Global GPP

When driven with the WFDEI-GPCC dataset (JULES-WFDEI-GPCC; Table 1), JULES simulates global GPP with an annual average of  $140 \text{ Pg C year}^{-1}$  (the combined GPP of all terrestrial ecosystems) over the 2001–2010 period (Figure 3c). This value is greater than that estimated by MODIS, FLUXNET-MTE and CARDAMOM with annual average global GPP estimated to be 130, 112 and  $114 \text{ Pg C year}^{-1}$ , respectively, for the same period (Figures 3a, b and d). The higher global GPP simulated by the JULES-WFDEI-GPCC driven simulations is greater than the MODIS, FLUXNET-MTE and CARDAMOM estimates by 25 %, 8 % and 23 % on average, respectively.

The difference in average annual global GPP between JULES-WFDEI-GPCC and MODIS (both at  $0.5^\circ \times 0.5^\circ$  spatial resolution) is greater ( $28 \text{ Pg C year}^{-1}$ ) than that between JULES-WFDEI-GPCC and FLUXNET-MTE ( $10 \text{ Pg C year}^{-1}$ ) and between JULES-WFDEI-GPCC and CARDAMOM ( $26 \text{ Pg C year}^{-1}$ ). This difference between the model simulated and observation-based GPP estimates is also shown in the zonal mean of the total annual JULES-WFDEI-GPCC, MODIS, FLUXNET-MTE and CARDAMOM GPP with the largest differences between datasets found in the tropics at  $10^\circ\text{S}$ – $10^\circ\text{N}$  and  $15^\circ\text{N}$ – $30^\circ\text{N}$  (Figure 3e).

In general, JULES simulates higher annual average global GPP than MODIS, FLUXNET-MTE and CARDAMOM with JULES GPP closer to FLUXNET-MTE GPP estimates.



### 3.3 Global and regional comparison of simulated GPP for various biomes

In addition to examining the ability of JULES to simulate global GPP (integrated across all ecosystem types), the total annual GPP for 2001–2010 was compared for various biomes (forests, grasslands and shrubs) at global and regional scales (Figure 4). This means that areas for model improvement can be identified at scales smaller than the global. When JULES was driven with WFDEI-GPCC (JULES-WFDEI-GPCC), JULES simulated total annual GPP to be 612 Pg C, 536 Pg C and 74 Pg C for forests, grasslands and shrubs, respectively with average annual GPP for forests, grasslands and shrubs being 61 Pg C year<sup>-1</sup>, 54 Pg C year<sup>-1</sup> and 7 Pg C year<sup>-1</sup>, respectively (Figure 4a). With the exception of shrubs, JULES overestimates total annual GPP by 31 %, 13 % and 22 % compared to MODIS, FLUXNET-MTE and CARDAMOM GPP, respectively, for forests and by 27 %, 10 % and 31 % compared to MODIS, FLUXNET-MTE and CARDAMOM GPP, respectively, for grasslands (Figure 4a). Differences between JULES, MODIS, FLUXNET-MTE and CARDAMOM GPP for shrubs are small with total annual GPP ranging within 68–78 Pg C (Figure 4a).

The differences in total annual GPP at the global scale is mainly due to differences between model simulated (JULES and CARDAMOM) and the observation-based estimates (MODIS and FLUXNET-MTE) in the tropics (30°S–30°N) (Figure 4b). In the tropics, JULES simulates total annual GPP to be 501 Pg C, 387 Pg C and 52 Pg C for forests, grasslands and shrubs, respectively, for the 2001–2010 period. JULES overestimates total annual GPP by 41 %, 20 % and 35 % compared to MODIS, FLUXNET-MTE and CARDAMOM GPP, respectively, for forests and by 47 %, 24 % and 54 % compared to MODIS, FLUXNET-MTE and CARDAMOM GPP, respectively, for grasslands in the tropical regions (Figure 4b). Differences between model simulated and observation-based estimates of GPP are small in the tropics for shrubs with total annual GPP ranging from 44–52 Pg C (Figure 4b). In the extratropics (30°N–90°N and 30°S–90°S), differences between model and observed GPP are small with total annual GPP for forests, grasslands and shrubs found to be 111–128 Pg C, 148–178 Pg C and 21–28 Pg C, respectively (Figure 4c).

Total annual GPP at the regional scale was examined by dividing the land area into seven regions (Figure 1; Table 2). The tropical regions (30°S–30°N) have been further divided up into three regions; Central and South America, Africa and South and South-East Asia. The extratropics (30°N–90°N and 30°S–90°S) have been divided into four regions; Europe, Northern Asia, North America and Greenland and the extratropical Southern Hemisphere. By normalising the FLUXNET-MTE and MODIS GPP by JULES GPP for these three regions (Figure 6), it is easier to see the differences between the model simulated and observation-based estimates of GPP. The dashed line at  $y=1$  for each of the seven regions in Figure 6 represents where model and observation-based total annual GPP match.

In the extratropics, differences between JULES, MODIS, FLUXNET-MTE and CARDAMOM GPP are small with total annual GPP ranging from 16–20 Pg C and 42–55 Pg C for forests and grasslands, respectively, in Europe, 42–53 Pg C and 46–59 Pg C for forests and grasslands, respectively, in Northern Asia, 6–8 Pg C and 13–18 Pg C for forests and grasslands, respectively, in the Extratropical Southern Hemisphere and 42–49 Pg C and 37–44 Pg C for forests and grasslands, respectively, in North America and Greenland (Figures 6a, b, d and g, respectively).

In general, JULES overestimates GPP in the tropics, but is able to simulate it in the extratropics.



### 3.4 Sensitivity to spatial resolution

When simulations of GPP were performed at lower spatial resolutions (JULES-WFDEI-GPCC-1degree and JULES-WFDEI-GPCC-2degree; Table 1), the average annual global GPP at  $0.5^\circ \times 0.5^\circ$ ,  $1^\circ \times 1^\circ$  and  $2^\circ \times 2^\circ$  spatial resolutions was 140 Pg C year<sup>-1</sup>, 141 Pg C year<sup>-1</sup> and 142 Pg C year<sup>-1</sup>, respectively. The percentage differences between JULES and the observation-based GPP estimates (MODIS and FLUXNET-MTE) at the various spatial resolutions are approximately equal with JULES differing from MODIS and FLUXNET-MTE by 8 % and 25 %, respectively, at  $0.5^\circ \times 0.5^\circ$  spatial resolution, by 8 % and 26 %, respectively, at  $1^\circ \times 1^\circ$  resolution and by 9 % and 26 %, respectively, at  $2^\circ \times 2^\circ$  resolution.

The zonal mean of modelled total annual GPP at various spatial resolutions are approximately equal (Figure 5). This insensitivity to spatial resolution is also found at regional scales (Figure 6). This insensitivity to spatial resolution is a useful result since it means that model simulations can be performed at  $2^\circ \times 2^\circ$  resolution with little difference to model output from the simulations at  $0.5^\circ \times 0.5^\circ$  and due to the lower computational cost, model run times (at  $2^\circ \times 2^\circ$  resolution) are short (approximately 16× faster than the  $0.5^\circ \times 0.5^\circ$  resolution simulations).

When simulating GPP at global and regional scales, there was little impact from varying spatial resolution ( $0.5^\circ \times 0.5^\circ$ ,  $1^\circ \times 1^\circ$  and  $2^\circ \times 2^\circ$ ).

### 3.5 Sensitivity to meteorological data set

When driven with the PRINCETON dataset (JULES-PRINCETON; Table 1), JULES simulates global GPP with an annual average of 144 Pg C year<sup>-1</sup> for the 2001–2010 period. As observed when driving JULES with the WFDEI-GPCC dataset (Figure 3), JULES-PRINCETON simulates higher global GPP than MODIS, FLUXNET-MTE and CARDAMOM at  $1^\circ \times 1^\circ$  spatial resolution by 29 %, 11 % and 26 % on average, respectively. This compares quite well to global GPP simulated by JULES when driven with the WFDEI-GPCC dataset, which had an annual average global GPP of 140 Pg C year<sup>-1</sup>. GPP simulated by JULES-WFDEI-GPCC was only higher than that of MODIS, FLUXNET-MTE (both at  $0.5^\circ \times 0.5^\circ$  spatial resolution) and CARDAMOM (at  $1^\circ \times 1^\circ$  resolution) by 25 %, 8 % and 23 % on average, respectively. The trend in zonal mean of total annual GPP simulated by the model (when driven with PRINCETON) is similar to that when driven with WFDEI-GPCC (at  $1^\circ \times 1^\circ$  spatial resolution) with differences being mostly in the tropics (Figure 5). The trend in differences between JULES-PRINCETON and JULES-WFDEI-GPCC-1degree and the observation-based estimates (MODIS and FLUXNET-MTE) is similar with model output from both simulations overestimating GPP in the tropics (Figure 5).

When driven with the WFDEI-CRU dataset, JULES simulates global GPP with an annual average of 141 Pg C year<sup>-1</sup> (the combined GPP of all terrestrial ecosystems) over 2001–2010. This is 1 Pg C year<sup>-1</sup> higher than that simulated when JULES is driven with WFDEI-GPCC (140 Pg C year<sup>-1</sup>). The small differences in global GPP can also be found at regional scales in both the tropical and extratropical regions (Figures 4b and c, respectively).

In general, when JULES is driven with the PRINCETON dataset, simulated global GPP was found to be higher than that simulated using WFDEI-GPCC by 3 Pg C year<sup>-1</sup> on average with the largest differences occurring in the tropics. There is



little difference in simulated GPP when using either WFDEI-GPCC or WFDEI-CRU (which differ only in the precipitation product used) to drive JULES.

## 4 Discussion

### 4.1 Can JULES capture interannual variability of GPP at the global scale? How do estimates of total annual GPP compare to those from observational datasets?

When JULES was driven with the WFDEI-GPCC dataset (at  $0.5^\circ \times 0.5^\circ$  spatial resolution), the model was able to capture interannual variability at the global scale (Figure 2c). This was also found when simulating GPP at lower spatial resolution ( $1^\circ \times 1^\circ$  and  $2^\circ \times 2^\circ$ ). At global scales, JULES estimates the annual average GPP to be  $140 \text{ Pg C year}^{-1}$  (combined GPP of all terrestrial ecosystems) over 2001–2010, which is greater than MODIS, FLUXNET-MTE and CARDAMOM GPP by 25 %, 8 % and 23 % on average, respectively (Figure 3). These differences are due to differences in GPP between JULES, the observation-based estimates (MODIS and FLUXNET-MTE) and CARDAMOM for forests and grasslands in the tropics (Figure 4b). In the extratropics, JULES was able to simulate GPP compared to MODIS, FLUXNET-MTE and CARDAMOM due to its phenology model and associated model parameters being designed for temperate regions.

The main difference between model simulated (JULES and CARDAMOM) and observation-based (MODIS and FLUXNET-MTE) estimates of GPP was found in the tropics with CARDAMOM GPP estimates being between the two observation-based datasets (Figure 3e). Photosynthesis is modelled differently in JULES and CARDAMOM. In JULES, leaf-level photosynthesis is calculated as the minimum of three limiting rates which is then scaled up to canopy level using the sum of the leaf-level fluxes in each canopy layer, scaled by leaf area (Clark et al., 2011). In CARDAMOM, GPP is calculated as a function of LAI, air temperature and radiation using the Aggregated Canopy Model (Williams et al., 1997, ACM). ACM is an emulator of the Soil Plant Atmosphere model (Williams et al., 1996, SPA).

When JULES was driven with WFDEI-GPCC (JULES-WFDEI-GPCC), it simulates lower GPP than MODIS, FLUXNET-MTE and CARDAMOM at  $15^\circ\text{N}$ – $30^\circ\text{N}$  (Figures 5 and 7). This difference in GPP was due to the incorrect simulation of GPP by JULES in Mexico (Figure 7). The total annual MODIS and FLUXNET-MTE GPP estimates for 2001–2010 are higher than that simulated by JULES by 1.0% and 6.7%, respectively, for Mexico, with CARDAMOM GPP estimates for the same period being lower than JULES GPP by 5.9%. One of the major vegetation types in Mexico is drought-deciduous plants (drought-deciduous plants lose their leaves during the dry season or periods of dryness as opposed to temperate deciduous plants which lose their leaves during periods of cold weather) and JULES does not contain this PFT.

Drought-deciduous plants can be found in the seasonally dry tropical forests of Mexico, Central America and northwestern South America. The implementation of drought-deciduous forest and shrub PFTs would help improve simulated GPP at latitudes  $15^\circ\text{N}$ – $30^\circ\text{N}$ . In JULES, phenology is updated once per day by multiplying the annual maximum LAI by a scaling factor, which is calculated using temperature-dependent leaf turnover rates. Leaf turnover rates are a function of surface air temperature and increase when the temperature drops below a certain value (this varies depending on the PFT). While this is suitable for deciduous broadleaf forests in temperate regions, such as Northern Europe, it will lead to inaccurate modelled



LAI for drought-deciduous forests. Instead of modifying modelled LAI using a temperature-derived scaling factor, the scaling factor could be calculated by using periods of dryness as the controlling factor.

#### 4.2 How do fluxes of GPP simulated by JULES compare for various biomes at the global and regional scales?

JULES (JULES-WFDEI-GPCC) simulates higher GPP than MODIS, FLUXNET-MTE and CARDAMOM at global scales  
5 and this was found to be due to higher GPP simulated by JULES for forests and grasslands in the tropics (Figure 4b). The total annual GPP for shrubs globally and in the tropics and extratropics are approximately equal (Figures 4a and c). JULES simulated average annual GPP to be 61 Pg C year<sup>-1</sup>, 54 Pg C year<sup>-1</sup> and 7 Pg C year<sup>-1</sup> for forests, grasslands and shrubs, respectively. Simulated GPP for forests is similar to that calculated by Beer et al. (2010) (sum of the values for tropical, temperate and boreal forests) with average annual GPP being 59 Pg C year<sup>-1</sup>. Since Beer et al. (2010) provides average annual GPP values  
10 for tropical savannahs and grasslands, temperate grasslands and shrublands and croplands, these are summed in order to obtain average annual global GPP for grasslands and shrubs 54.6 Pg C year<sup>-1</sup>, which is lower than the model simulated value of 61 Pg C year<sup>-1</sup>.

By further dividing the global land area into seven regions (Table 2), it was found that for the three tropical regions (Central and South America, Africa and South and South-East Asia), JULES overestimated total annual GPP for forests, grasslands  
15 and shrubs (Figures 6c, e, and f). The four extratropical regions (Europe, Northern Asia, Extratropical Southern Hemisphere and North America and Greenland) simulate similar GPP for JULES, MODIS and FLUXNET-MTE for the three biomes with shrubs in North America and Greenland, Northern Asia and the extratropical Southern Hemisphere being underestimated by JULES (Figures 6a, b, d and g).

#### 4.3 How sensitive are fluxes of GPP to the spatial resolution of the model?

20 When JULES was driven with the WFDEI-GPCC dataset at three different spatial resolutions (JULES-WFDEI-GPCC, JULES-WFDEI-GPCC-1degree and JULES-WFDEI-GPCC-2degree; Table 1), it was found that for GPP simulations, the model was insensitive to spatial resolution with average annual global GPP being 140Pg C year<sup>-1</sup>, 141 Pg C year<sup>-1</sup> and 142 Pg C year<sup>-1</sup> at 0.5° × 0.5°, 1° × 1° and 2° × 2° spatial resolutions, respectively. This trend was also observed in the zonal mean of total annual GPP (Figure 5). The insensitivity of the model to spatial resolution at the global scale was also observed at the regional  
25 scale when comparing simulated GPP fluxes for forests, grasslands and shrubs in the tropics and extratropics (Figure 6).

Little research has been performed on the effects of spatial resolution on JULES simulations (as well as other LSMs). Studies using atmospheric chemistry models have shown that the spatial resolution of the input meteorological data can affect model output (Ito et al., 2009; Pugh et al., 2013; Schaap et al., 2015). The results found here agree with those from Compton and Best (2011). Compton and Best (2011) showed that JULES was insensitive to spatial resolution when the WFD dataset  
30 was regridded from half-degree to 1-degree and 2-degree when simulating the terrestrial hydrological cycle. It was found that spatial resolution had little or no effect on simulations of global mean total evaporation and total runoff. However, the study showed that JULES was sensitive to temporal resolution when simulating the same hydrological components.



#### 4.4 Is the meteorological dataset used to drive the model important at the global scale?

When JULES was driven with the PRINCETON dataset at  $1^\circ \times 1^\circ$  spatial resolution (Table 1), the annual average global GPP was slightly higher by  $4 \text{ Pg C year}^{-1}$  than that simulated by JULES when driven with the WFDEI-GPCC dataset. In general, differences in GPP fluxes for model simulations driven using WFDEI-GPCC and PRINCETON are mainly in the tropics (at  
5  $5^\circ\text{N}$ – $5^\circ\text{S}$ ) with JULES-WFDEI-GPCC-1degree simulating higher GPP than JULES-PRINCETON and in the extratropics at  $30^\circ\text{N}$ – $60^\circ\text{N}$ , JULES-PRINCETON simulates slightly higher GPP (Figure 5).

Other studies have shown that the meteorological dataset used to drive LSMs is a large source of uncertainty in global land surface modelling (Hicke, 2005; Jung et al., 2007; Poulter et al., 2011). Different methods are used to create time series of global gridded climate data in order to drive LSMs and this can introduce uncertainty that can propagate through model  
10 simulations (Zhao et al., 2006). Even at the point scale, differences in simulated GPP were observed when driving JULES with the WFDEI-GPCC and PRINCETON datasets (Slevin et al., 2015). As in this study, it occurred in the tropics.

## 5 Conclusions

An evaluation of JULES was performed at global and regional scales with simulated GPP compared to global gridded ( $0.5^\circ \times 0.5^\circ$  spatial and monthly temporal resolution) estimates of GPP derived from upscaled FLUXNET observations (FLUXNET-  
15 MTE), satellite observations from the MODIS sensor and that produced by the CARDAMOM data assimilation framework. In general, it was found that JULES was able to capture interannual variability at the global scale. JULES simulated higher global GPP than FLUXNET-MTE, MODIS and CARDAMOM but at the regional scale, these differences were due to differences between model simulated and observation-based estimates in the tropics. In general, CARDAMOM was better at simulating GPP than JULES.

Differences in GPP between JULES and the benchmarking datasets (FLUXNET-MTE, MODIS and CARDAMOM) at  
20  $15^\circ\text{N}$ – $30^\circ\text{N}$  is due to higher FLUXNET-MTE, MODIS and CARDAMOM GPP for Mexico because of a lack of drought-deciduous PFTs in JULES. The inclusion of these PFTs would improve GPP simulations at latitude  $15^\circ\text{N}$ – $30^\circ\text{N}$  (mostly in Mexico). By dividing the global land area into seven regions, it was found that all three tropical regions (Central and South America, Africa and South and South-East Asia) contribute to model-observation differences at the global scale compared to  
25 FLUXNET-MTE and MODIS. The model is able simulate GPP estimates at the four extratropical regions (Europe, Northern Asia, North America and Greenland and the extratropical Southern Hemisphere).

Improved GPP simulations in the tropics can be attained with the introduction of more PFT classes and their associated model parameters. In this study, the version of JULES used was 3.4.1. In this version, each model grid box is composed of nine different surface types and five of these are PFTs. Since model version 4.2, each JULES gridbox contains nine PFTs (tropical  
30 broadleaf evergreen, temperate broadleaf evergreen, broadleaf deciduous, needleleaf evergreen, needleleaf deciduous, C3, C4, evergreen shrub, deciduous shrub) Harper et al. (2016). In addition to these PFTs, a phenology model which can simulate



LAI in both temperate and tropical regions, would help to reduce differences between model simulated and observation-based estimates of GPP in the tropics.

When JULES was driven at the global and regional scale with the WFDEI-GPCC dataset at various spatial resolutions ( $0.5^\circ \times 0.5^\circ$ ,  $1^\circ \times 1^\circ$  and  $2^\circ \times 2^\circ$ ), it was found that the model was insensitive to spatial resolution. Similar results were shown by Compton and Best (2011) when simulating components of the terrestrial hydrological cycle. Differences between high ( $0.5^\circ \times 0.5^\circ$ ) and low ( $2^\circ \times 2^\circ$ ) spatial resolution simulations of GPP are very similar. This means that low spatial resolution model simulations at these scales can be performed in place of high resolution when simulating GPP and results in shorter model run times.

The meteorological dataset used to drive LSMs at the global scale is an important source of model uncertainty (Poulter et al., 2011). By using a different meteorological dataset (PRINCETON) to drive the model, it was found that simulated GPP was similar to that when the model was driven with the WFDEI-GPCC dataset (at  $1^\circ \times 1^\circ$  spatial resolution) with exceptions to this being in the tropics and the northern extratropics. These differences are due to biases in the downward radiation fluxes and surface air temperature in the meteorological data. Differences in precipitation, and hence soil moisture stress, did not play a role in differences between the two model simulations. When JULES was driven with the WFDEI-CRU dataset instead of WFDEI-GPCC, differences in simulated GPP were very small.

In general, differences between JULES GPP and the benchmarking datasets (FLUXNET-MTE, MODIS and CARDAMOM) occur mostly in the tropics with differences at  $15^\circ\text{N}$ – $30^\circ\text{N}$  due a lack of drought-deciduous PFTs in JULES. There was little difference in JULES GPP when the model was driven with different meteorological datasets. Finally, at large scales JULES GPP was insensitive to spatial resolution.

## 6 Code and/or data availability

The outputs from the JULES model simulations reported in this paper have been deposited online at DataShare, the University of Edinburgh's digital repository of multidisciplinary research datasets, <http://dx.doi.org/10.7488/ds/1461>.

*Author contributions.* D.S., S.F.B.T. and M.W. designed the research. D.S., J.-F.E. and A.A.B performed the model simulations. D.S., S.F.B.T. and M.W. analysed the data. D.S. prepared the manuscript with contributions from all co-authors.

*Acknowledgements.* I would like to gratefully acknowledge the School of GeoSciences, University of Edinburgh, which provided me with a School Scholarship that allowed me to pursue my PhD research, during which this work was carried out. Part of this work was conducted at the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. This work would not have been possible without help from the following people. Rich Ellis at the Centre for Ecology and Hydrology (CEH) provided the soil and vegetation ancillary files and advice on modifying the namelist files in order to set up the model simulations to run on a 2d grid. Doug Clark (also at CEH) provided advice on model spin-up at global scales. Information on the snow schemes used by JULES and advice on which one to use at global scales was provided by Richard Essery. Graham Weedon provided invaluable advice on using the WFDEI meteorological dataset.



## References

- 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.: Spatio-temporal patterns of terrestrial gross primary production: A review, *Reviews of Geophysics*, doi:10.1002/2015RG000483, 2015.
- 5 Arora, V. K., Boer, G. J., Friedlingstein, P., Eby, M., Jones, C. D., Christian, J. R., Bonan, G., Bopp, L., Brovkin, V., Cadule, P., Hajima, T., Ilyina, T., Lindsay, K., Tjiputra, J. F., and Wu, T.: Carbon–Concentration and Carbon–Climate Feedbacks in CMIP5 Earth System Models, *Journal of Climate*, 26, 5289–5314, doi:10.1175/JCLI-D-12-00494.1, 2013.
- Battin, T. J., Luysaert, S., Kaplan, L. A., Aufdenkampe, A. K., Richter, A., and Tranvik, L. J.: The boundless carbon cycle, *Nature Geoscience*, 2, 598–600, doi:10.1038/ngeo618, 2009.
- 10 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., Luysaert, S., Margolis, H., Oleson, K. W., Rouspard, 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, doi:10.1126/science.1184984, 2010.
- Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., 15 Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), Model description–Part 1: Energy and water fluxes, *Geoscientific Model Development*, 4, 677–699, doi:10.5194/gmd-4-677-2011, 2011.
- Bloom, A. A. and Williams, M.: Constraining ecosystem carbon dynamics in a data-limited world: integrating ecological "common sense" in a model–data fusion framework, *Biogeosciences*, 12, 1299–1315, doi:10.5194/bg-12-1299-2015, 2015.
- 20 Bloom, A. A., Exbrayat, J.-F., van der Velde, I. R., Feng, L., and Williams, M.: The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times, *Proceedings of the National Academy of Sciences*, 113, 1285–1290, doi:10.1073/pnas.1515160113, 2016.
- Blyth, E., Gash, J., Lloyd, A., Pryor, M., Weedon, G. P., and Shuttleworth, J.: Evaluating the JULES Land Surface Model Energy Fluxes Using FLUXNET Data, *Journal of Hydrometeorology*, 11, 509–519, doi:10.1175/2009JHM1183.1, 2010.
- 25 Blyth, E., Clark, D. B., Ellis, R., Huntingford, C., Los, S., Pryor, M., Best, M., and Sitch, S.: A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale, *Geoscientific Model Development*, 4, 255–269, doi:10.5194/gmd-4-255-2011, 2011.
- Bonan, G. B.: Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests, *Science*, 320, 1444–1449, doi:10.1126/science.1155121, 2008.
- 30 Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., Lawrence, D. M., and Swenson, S. C.: Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data, *Journal of Geophysical Research*, 116, G02 014, doi:10.1029/2010JG001593, 2011.
- Burke, E. J., Dankers, R., Jones, C. D., and Wiltshire, A. J.: A retrospective analysis of pan Arctic permafrost using the JULES land surface model, *Climate Dynamics*, 41, 1025–1038, doi:10.1007/s00382-012-1648-x, 2013.
- 35 Canadell, J. G., Le Quéré, C., Raupach, M. R., Field, C. B., Buitenhuis, E. T., Ciais, P., Conway, T. J., Gillett, N. P., Houghton, R. A., and Marland, G.: Contributions to accelerating atmospheric CO<sub>2</sub> growth from economic activity, carbon intensity, and efficiency of natural sinks,



- Proceedings of the National Academy of Sciences of the United States of America, 104, 18 866–18 870, doi:10.1073/pnas.0702737104, 2007.
- Chadburn, S., Burke, E., Essery, R., Boike, J., Langer, M., Heikenfeld, M., Cox, P., and Friedlingstein, P.: An improved representation of physical permafrost dynamics in the JULES land-surface model, *Geoscientific Model Development*, 8, 1493–1508, doi:10.5194/gmd-8-1493-2015, 2015.
- 5 Chapin III, F. S., Matson, P. A., and Vitousek, P.: *Principles of Terrestrial Ecosystem Ecology*, Springer-Verlag New York, doi:10.1007/978-1-4419-9504-9, 2012.
- Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., Huntingford, C., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description–Part 2: Carbon fluxes and vegetation dynamics, *Geoscientific Model Development*, 4, 701–722, doi:10.5194/gmd-4-701-2011, 2011.
- 10 Collatz, G. J., Ball, J. T., Grivet, C., and Berry, J. A.: Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer, *Agricultural and Forest Meteorology*, 54, 107–136, doi:10.1016/0168-1923(91)90002-8, 1991.
- Collatz, G. J., Ribas-Carbo, M., and Berry, J. A.: Coupled photosynthesis-stomatal conductance model for leaves of C<sub>4</sub> plants, *Functional Plant Biology*, 19, 519–538, doi:10.1071/PP9920519, 1992.
- 15 Compton, E. and Best, M.: Impact of spatial and temporal resolution on modelled terrestrial hydrological cycle components, Tech. Rep. 44, WATCH Technical Report, 19 pp [Available at <http://www.eu-watch.org/>], 2011.
- Cox, P. and Jones, C.: Illuminating the Modern Dance of Climate and CO<sub>2</sub>, *Science*, 321, 1642–1644, doi:10.1126/science.1158907, 2008.
- Cox, P. M.: Description of the "TRIFFID" Dynamic Global Vegetation Model, Tech. Rep. 24, Hadley Centre, Met Office, London Road, Bracknell, Berkshire, RG12 2SY, UK, 2001.
- 20 Cox, P. M., Huntingford, C., and Harding, R. J.: A canopy conductance and photosynthesis model for use in a GCM land surface scheme, *Journal of Hydrology*, 212–213, 79–94, doi:10.1016/S0022-1694(98)00203-0, 1998.
- Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R., and Smith, J.: The impact of new land surface physics on the GCM simulation of climate and climate sensitivity, *Climate Dynamics*, 15, 183–203, doi:10.1007/s003820050276, 1999.
- 25 Dalmonech, D., Zaehle, S., Schürmann, G. J., Brovkin, V., Reick, C., and Schnur, R.: Separation of the Effects of Land and Climate Model Errors on Simulated Contemporary Land Carbon Cycle Trends in the MPI Earth System Model version 1, *Journal of Climate*, 28, 272–291, doi:10.1175/JCLI-D-13-00593.1, 2014.
- Deardorff, J. W.: Efficient Prediction of Ground Surface Temperature and Moisture, With Inclusion of a Layer of Vegetation, *Journal of Geophysical Research*, 83, 1889–1903, doi:10.1029/JC083iC04p01889, 1978.
- 30 Foley, J. A., Prentice, I. C., Ramankutty, N., Levis, S., Pollard, D., Sitch, S., and Haxeltine, A.: An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics, *Global Biogeochemical Cycles*, 10, 603–628, doi:10.1029/96GB02692, 1996.
- Friedlingstein, P., Cox, P., Betts, R., Bopp, L., Von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., et al.: Climate-Carbon Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison, *Journal of Climate*, 19, 3337–3353, doi:10.1175/JCLI3800.1, 2006.
- 35 Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., and Knutti, R.: Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks, *Journal of Climate*, 27, 511–526, doi:10.1175/JCLI-D-12-00579.1, 2014.



- Galbraith, D., Levy, P. E., Sitch, S., Huntingford, C., Cox, P., Williams, M., and Meir, P.: Multiple mechanisms of Amazonian forest biomass losses in three dynamic global vegetation models under climate change, *New Phytologist*, 187, 647–665, doi:10.1111/j.1469-8137.2010.03350.x, 2010.
- Giglio, L., Randerson, J. T., and van der Werf, G. R.: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), *Journal of Geophysical Research: Biogeosciences*, 118, 317–328, doi:10.1002/jgrg.20042, 2013.
- Harding, R., Best, M., Blyth, E., Hagemann, S., Kabat, P., Tallaksen, L. M., Warnaars, T., Wiberg, D., Weedon, G. P., van Lanen, H., Ludwig, F., and Haddeland, I.: WATCH: Current Knowledge of the Terrestrial Global Water Cycle, *Journal of Hydrometeorology*, 12, 1149–1156, doi:10.1175/JHM-D-11-024.1, 2011.
- Harper, A. B., Cox, P. M., Friedlingstein, P., Wiltshire, A. J., Jones, C. D., Sitch, S., Mercado, L. M., Groenendijk, M., Robertson, E., Kattge, J., Bönisch, G., Atkin, O. K., Bahn, M., Cornelissen, J., Niinemets, U., Onipchenko, V., Peñuelas, J., Poorter, L., Reich, P. B., Soudzilovskaia, N. A., and Bodegom, P. V.: Improved representation of plant functional types and physiology in the Joint UK Land Environment Simulator (JULES v4.2) using plant trait information, *Geoscientific Model Development*, 9, 2415–2440, doi:10.5194/gmd-9-2415-2016, 2016.
- Hicke, J. A.: NCEP and GISS solar radiation data sets available for ecosystem modeling: Description, differences, and impacts on net primary production, *Global Biogeochemical Cycles*, 19, GB2006, doi:10.1029/2004GB002391, 2005.
- Hiederer, R. and Köchy, M.: *Global Soil Organic Carbon Estimates and the Harmonized World Soil Database*, Tech. Rep. EUR 25225 EN, Publications Office of the European Union, Luxembourg, 2011.
- Ito, A., Sillman, S., and Penner, J. E.: Global chemical transport model study of ozone response to changes in chemical kinetics and biogenic volatile organic compounds emissions due to increasing temperatures: Sensitivities to isoprene nitrate chemistry and grid resolution, *Journal of Geophysical Research: Atmospheres* (1984–2012), 114, D09 301, doi:10.1029/2008JD011254, 2009.
- Jacobs, C. M. J.: *Direct impact of atmospheric CO<sub>2</sub> enrichment on regional transpiration*, Ph.D. thesis, Wageningen Agricultural University, 1994.
- Janssens, I. A., Freibauer, A., Ciais, P., Smith, P., Nabuurs, G.-J., Folberth, G., Schlamadinger, B., Hutjes, R. W. A., Ceulemans, R., Schulze, E.-D., Valentini, R., and Dolman, A. J.: Europe's Terrestrial Biosphere Absorbs 7 to 12% of European Anthropogenic CO<sub>2</sub> Emissions, *Science*, 300, 1538–1542, doi:10.1126/science.1083592, 2003.
- Jung, M., Vetter, M., Herold, M., Churkina, G., Reichstein, M., Zaehle, S., Ciais, P., Viovy, N., Bondeau, A., Chen, Y., Trusilova, K., Feser, F., and Heimann, M.: Uncertainties of modeling gross primary productivity over Europe: A systematic study on the effects of using different drivers and terrestrial biosphere models, *Global Biogeochemical Cycles*, 21, GB4021, doi:10.1029/2006GB002915, 2007.
- Jung, M., Reichstein, M., and Bondeau, A.: Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model, *Biogeosciences*, 6, 2001–2013, doi:10.5194/bg-6-2001-2009, 2009.
- Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., de Jeu, R., Dolman, A. J., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B. E., Montagnani, L., Mu, Q., Mueller, B., Oleson, K., Papale, D., Richardson, A. D., Rouspard, O., Running, S., Tomelleri, E., Viovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S., and Zhang, K.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, *Nature*, 467, 951–954, doi:10.1038/nature09396, 2010.
- Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F., and Williams, C.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and



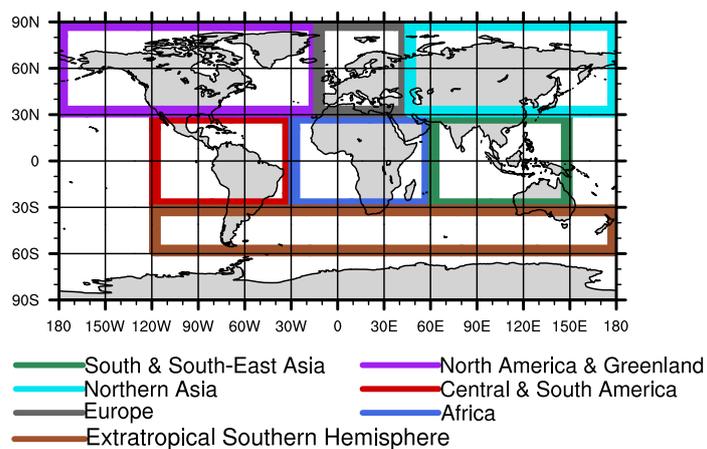
- sensible heat derived from eddy covariance, satellite, and meteorological observations, *Journal of Geophysical Research*, 116, G00J07, doi:10.1029/2010JG001566, 2011.
- Lasslop, G., Reichstein, M., Papale, D., Richardson, A. D., Arneeth, A., Barr, A., Stoy, P., and Wohlfahrt, G.: Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation, *Global Change Biology*, 16, 187–208, doi:10.1111/j.1365-2486.2009.02041.x, 2010.
- 5 Le Quéré, C., Raupach, M. R., Canadell, J. G., and Marland, G.: Trends in the sources and sinks of carbon dioxide, *Nature Geoscience*, 2, 831–836, doi:10.1038/ngeo689, 2009.
- Le Quéré, C., Andres, R. J., Boden, T., Conway, T., Houghton, R. A., House, J. I., Marland, G., Peters, G. P., van der Werf, G. R., Ahlström, A., Andrew, R. M., Bopp, L., Canadell, J. G., Ciais, P., Doney, S. C., Enright, C., Friedlingstein, P., Huntingford, C., Jain, A. K., Jourdain, C., Kato, E., Keeling, R. F., Klein Goldewijk, K., Levis, S., Levy, P., Lomas, M., Poulter, B., Raupach, M. R., Schwinger, J., Sitch, S., Stocker, B. D., Viovy, N., Zaehle, S., and Zeng, N.: The global carbon budget 1959–2011, *Earth System Science Data*, 5, 165–185, doi:10.5194/essd-5-165-2013, 2013.
- 10 Lei, H., Huang, M., Leung, L. R., Yang, D., Shi, X., Mao, J., Hayes, D. J., Schwalm, C. R., Wei, Y., and Liu, S.: Sensitivity of global terrestrial gross primary production to hydrologic states simulated by the Community Land Model using two runoff parameterizations, *Journal of Advances in Modeling Earth Systems*, 6, 658–679, doi:10.1002/2013MS000252, 2014.
- 15 Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., and Merchant, J. W.: Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, *International Journal of Remote Sensing*, 21, 1303–1330, doi:10.1080/014311600210191, 2000.
- Manabe, S.: Climate and the ocean circulation: 1, the atmospheric circulation and the hydrology of the Earth's surface, *Monthly Weather Review*, 97, 739–805, doi:10.1175/1520-0493(1969)097<0739:CATOC>2.3.CO;2, 1969.
- 20 Ménard, C. B., Ikonen, J., Rautiainen, K., Aurela, M., Arslan, A. N., and Pulliainen, J.: Effects of Meteorological and Ancillary Data, Temporal Averaging, and Evaluation Methods on Model Performance and Uncertainty in a Land Surface Model, *Journal of Hydrometeorology*, 16, 2559–2576, doi:10.1175/JHM-D-15-0013.1, 2015.
- Nachtergaele, F., van Velthuisen, H., Verelst, L., Wiberg, D., Batjes, N., Dijkshoorn, K., van Engelen, V., Fischer, G., Jones, A., Montanarella, L., Petri, M., Prieler, S., Teixeira, E., and Shi, X.: Harmonized World Soil Database v1.2, Tech. rep., International Institute for Applied Systems Analysis (IIASA), Food and Agriculture Organization of the United Nations (FAO), 2012.
- 25 Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., Phillips, O. L., Shvidenko, A., Lewis, S. L., Canadell, J. G., Ciais, P., Jackson, R. B., Pacala, S. W., McGuire, A. D., Piao, S., Rautiainen, A., Sitch, S., and Hayes, D.: A Large and Persistent Carbon Sink in the World's Forests, *Science*, 333, 988–993, doi:10.1126/science.1201609, 2011.
- 30 Pielke, R. A., Avissar, R., Raupach, M., Dolman, A. J., Zeng, X., and Denning, A. S.: Interactions between the atmosphere and terrestrial ecosystems: influence on weather and climate, *Global Change Biology*, 4, 461–475, doi:10.1046/j.1365-2486.1998.t01-1-00176.x, 1998.
- Pitman, A. J.: The evolution of, and revolution in, land surface schemes designed for climate models, *International Journal of Climatology*, 23, 479–510, doi:10.1002/joc.893, 2003.
- Poulter, B., Frank, D. C., Hodson, E. L., and Zimmermann, N. E.: Impacts of land cover and climate data selection on understanding terrestrial carbon dynamics and the CO<sub>2</sub> airborne fraction, *Biogeosciences*, 8, 2027–2036, doi:10.5194/bg-8-2027-2011, 2011.
- 35 Pugh, T. A. M., Ashworth, K., Wild, O., and Hewitt, C. N.: Effects of the spatial resolution of climate data on estimates of biogenic isoprene emissions, *Atmospheric Environment*, 70, 1–6, doi:10.1016/j.atmosenv.2013.01.001, 2013.



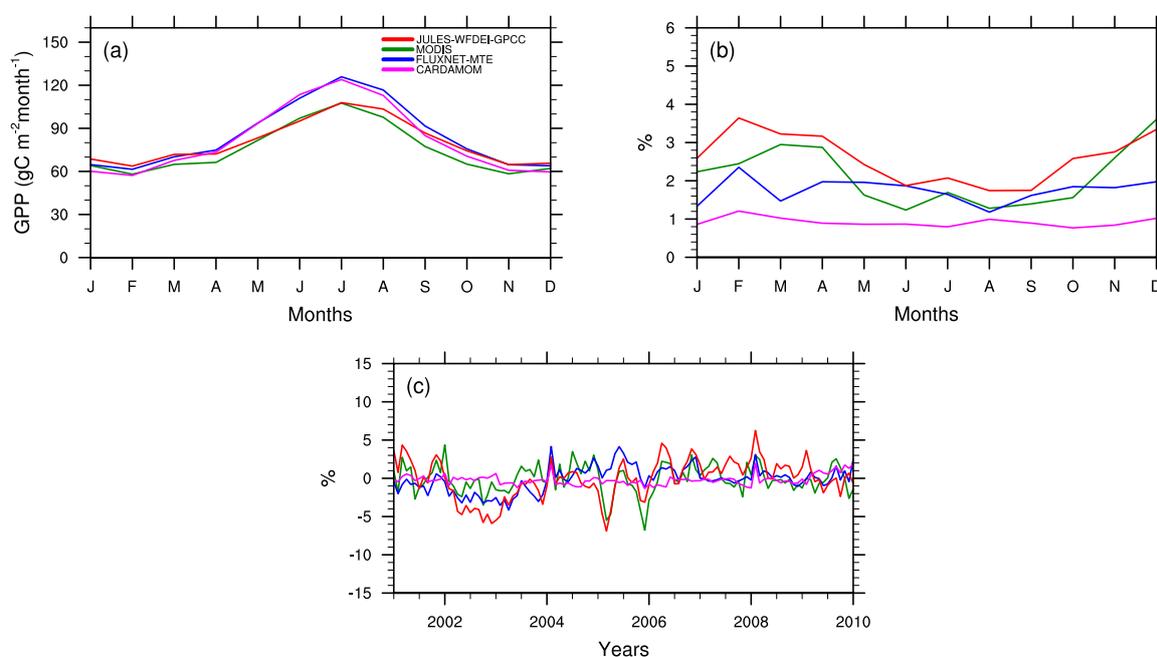
- Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, K., Ilvesniemi, H. and Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., and Valentini, R.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved  
5 algorithm, *Global Change Biology*, 11, 1424–1439, doi:10.1111/j.1365-2486.2005.001002.x, 2005.
- Running, S. W., Thornton, P. E., Nemani, R., and Glassy, J. M.: Global Terrestrial Gross and Net Primary Productivity from the Earth Observing System, in: *Methods in Ecosystem Science*, pp. 44–57, Springer, 2000.
- Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S., Petrova, S., White, L., Silman, M., and Morel, A.: Benchmark map of forest carbon stocks in tropical regions across three continents,  
10 *Proceedings of the National Academy of Sciences*, 108, 9899–9904, doi:10.1073/pnas.1019576108, 2011.
- Schaap, M., Cuvelier, C., Hendriks, C., Bessagnet, B., Baldasano, J. M., Colette, A., Thunis, P., Karam, D., Fagerli, H., Graff, A., Kranenburg, R., Nyiri, A., Pay, M. T., Rouil, L., Schulz, M., Simpson, D., Stern, R., Terrenoire, E., and Wind, P.: Performance of European chemistry transport models as function of horizontal resolution, *Atmospheric Environment*, 112, 90–105, doi:10.1016/j.atmosenv.2015.04.003, 2015.
- Schaefer, K., Schwalm, C. R., Williams, C., Arain, M. A., Barr, A., Chen, J. M., Davis, K. J., Dimitrov, D., Hilton, T. W., Hollinger, D. Y.,  
15 Humphreys, E., Poulter, B., Raczka, B. M., Richardson, A. D., Sahoo, A., Thornton, P., Vargas, R. and Verbeeck, H., Anderson, R., Baker, I., Black, T. A., Bolstad, P., Chen, J., Curtis, P. S., Desai, A. R., Dietze, M., Dragoni, D., Gough, C., Grant, R. F., Gu, L., Jain, A., Kucharik, C., Law, B., Liu, S., Lokipitiya, E., Margolis, H. A., Matamala, R., McCaughey, J. H., Monson, R., Munger, J. W., Oechel, W., Peng, C., Price, D. T., Ricciuto, D., Riley, W. J., Roulet, N., Tian, H., Tonitto, C., Torn, M., Weng, E., and Zhou, X.: A model-data comparison of gross primary productivity: Results from the North American Carbon Program site synthesis, *Journal of Geophysical Research*, 117,  
20 G03 010, doi:10.1029/2012JG001960, 2012.
- Sellers, P., Mintz, Y., Sud, Y., and Dalcher, A.: A Simple Biosphere Model (SiB) for Use within General Circulation Models, *Journal of the Atmospheric Sciences*, 43, 505–531, doi:10.1175/1520-0469(1986)043<0505:ASBMFU>2.0.CO;2, 1986.
- Seneviratne, S. I. and Stöckli, R.: Climate Variability and Extremes during the Past 100 Years , chap. *The Role of Land-Atmosphere Interactions for Climate Variability in Europe*, pp. 179–193, Springer, 2008.
- 25 Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling, *Journal of Climate*, 19, 3088–3111, doi:10.1175/JCLI3790.1, 2006.
- Sitch, S., Friedlingstein, P., Gruber, N., Jones, S. D., Murray-Tortarolo, G., Ahlström, A., Doney, S. C., Graven, H., Heinze, C., Huntingford, C., Levis, S., Levy, P. E., Lomas, M., Poulter, B., Viovy, N., Zaehle, S., Zeng, N., Arneth, A., Bonan, G., Bopp, L., Canadell, J. G., Chevallier, F., Ciais, P., Ellis, R., Gloor, M., Peylin, P., Piao, S. L., Le Quéré, C., Smith, B., Zhu, Z., and Myneni, R.: Recent trends and  
30 drivers of regional sources and sinks of carbon dioxide, *Biogeosciences*, 12, 653–679, doi:10.5194/bg-12-653-2015, 2015.
- Slevin, D., Tett, S. F. B., and Williams, M.: Multi-site evaluation of the JULES land surface model using global and local data, *Geoscientific Model Development*, 8, 295–316, doi:10.5194/gmd-8-295-2015, 2015.
- Tian, H., Lu, C., Ciais, P., Michalak, A. M., Canadell, J. G., Saikawa, E., Huntzinger, D. N., Gurney, K. R., Sitch, S., Zhang, B., Yang, J., Bousquet, P., Bruhwiler, L., Chen, G., Dlugokencky, E., Friedlingstein, P., Melillo, J., Pan, S., Poulter, B., Prinn, R., Saunio, M.,  
35 Schwalm, C. R., and Wofsy, S. C.: The terrestrial biosphere as a net source of greenhouse gases to the atmosphere, *Nature*, 531, 225–228, doi:10.1038/nature16946, 2016.
- van den Hurk, B., Best, M., Dirmeyer, P., Pitman, A., Polcher, J., and Santanello, J.: Acceleration of land surface model development over a decade of GLASS, *Bulletin of the American Meteorological Society*, 92, 1593–1600, doi:10.1175/BAMS-D-11-00007.1, 2011.



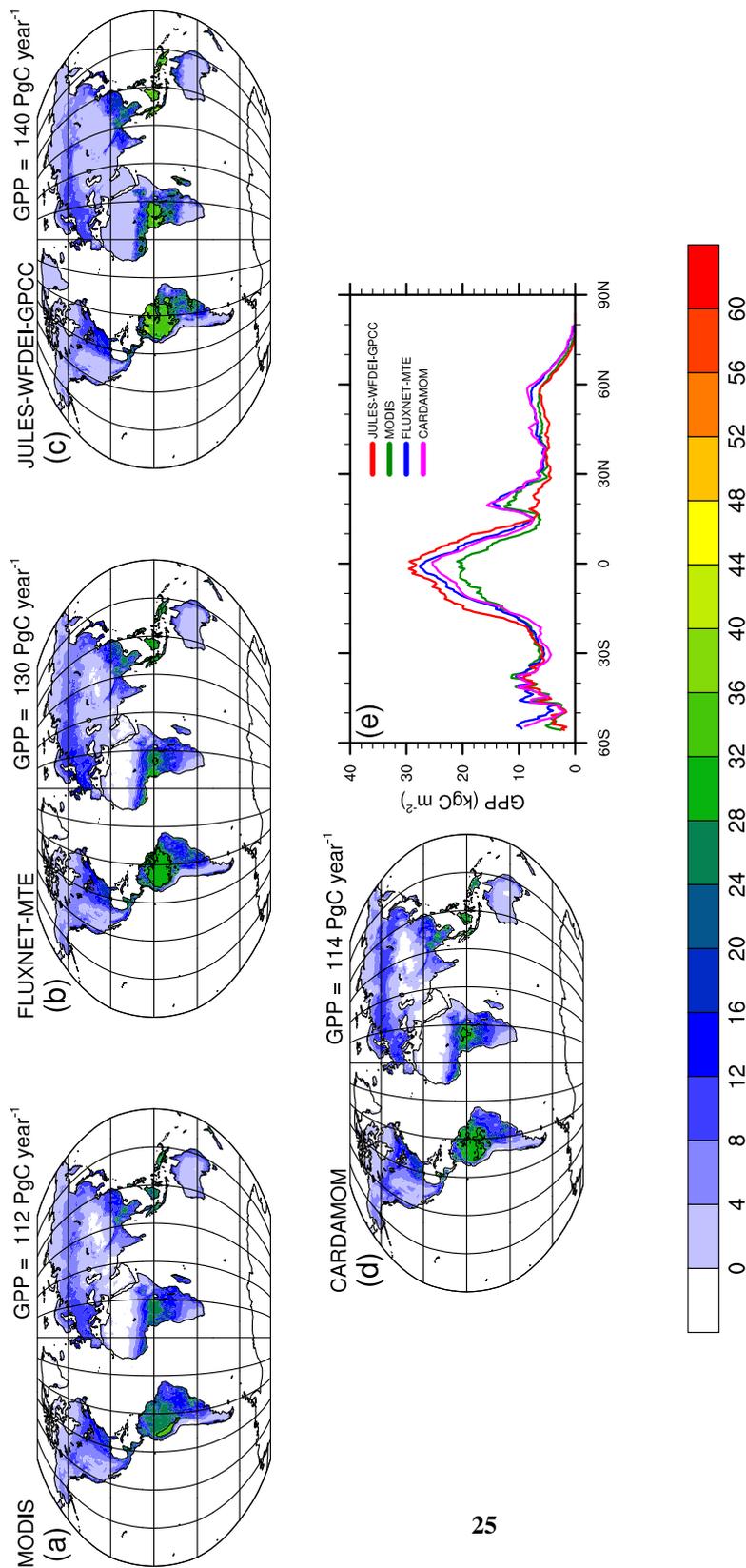
- Walters, D. N., Williams, K. D., Boutle, I. A., Bushell, A. C., Edwards, J. M., Field, P. R., Lock, A. P., Morcrette, C. J., Stratton, R. A., Wilkinson, J. M., Willett, M. R., Bellouin, N., Bodas-Salcedo, A., Brooks, M. E., Copsey, D., Earnshaw, P. D., Hardiman, S. C., Harris, C. M., Levine, R. C., MacLachlan, C., Manners, J. C., Martin, G. M., Milton, S. F., Palmer, M. D., Roberts, M. J., Rodríguez, J. M., Tennant, W. J., and Vidale, P. L.: The Met Office Unified Model Global Atmosphere 4.0 and JULES Global Land 4.0 configurations, 5 Geoscientific Model Development, 7, 361–386, doi:10.5194/gmd-7-361-2014, 2014.
- Weedon, G. P., Gomes, S., Viterbo, P., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: The WATCH Forcing Data 1958–2001: A meteorological forcing dataset for land surface- and hydrological-models, Tech. Rep. 22, WATCH Technical Report, 41 pp [Available at <http://www.eu-watch.org/>], 2010.
- Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: 10 Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century, *Journal of Hydrometeorology*, 12, 823–848, doi:10.1175/2011JHM1369.1, 2011.
- Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, *Water Resources Research*, 50, 7505–7514, doi:10.1002/2014WR015638, 2014.
- 15 Williams, M., Rastetter, E. B., Fernandes, D. N., Goulden, M. L., Wofsy, S. C., Shaver, G. R., Melillo, J. M., Munger, J. W., Fan, S.-M., and Nadelhoffer, K. J.: Modelling the soil-plant-atmosphere continuum in a Quercus–Acer stand at Harvard Forest: the regulation of stomatal conductance by light, nitrogen and soil/plant hydraulic properties, *Plant, Cell & Environment*, 19, 911–927, doi:10.1111/j.1365-3040.1996.tb00456.x, 1996.
- Williams, M., Rastetter, E. B., Fernandes, D. N., Goulden, M. L., Shaver, G. R., and Johnson, L. C.: Predicting Gross Primary Productivity 20 in Terrestrial Ecosystems, *Ecological Applications*, 7, 882–894, doi:10.1890/1051-0761(1997)007[0882:PGPPIT]2.0.CO;2, 1997.
- Yang, W., Tan, B., Huang, D., Rautiainen, M., Shabanov, N. V., Wang, Y., Privette, J. L., Huemmrich, K. F., Fensholt, R., Sandholt, I., et al.: MODIS leaf area index products: From validation to algorithm improvement, *Geoscience and Remote Sensing, IEEE Transactions on*, 44, 1885–1898, doi:10.1109/TGRS.2006.871215, 2006.
- Zhao, M. and Running, S. W.: Drought-induced reduction in global terrestrial net primary production from 2000 through 2009, *Science*, 329, 25 940–943, doi:10.1126/science.1192666, 2010.
- 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, doi:10.1016/j.rse.2004.12.011, 2005.
- Zhao, M., Running, S. W., and Nemani, R. R.: Sensitivity of Moderate Resolution Imaging Spectroradiometer (MODIS) terrestrial primary production to the accuracy of meteorological reanalyses, *Journal of Geophysical Research: Biogeosciences*, 111, 30 doi:10.1029/2004JG000004, 2006.



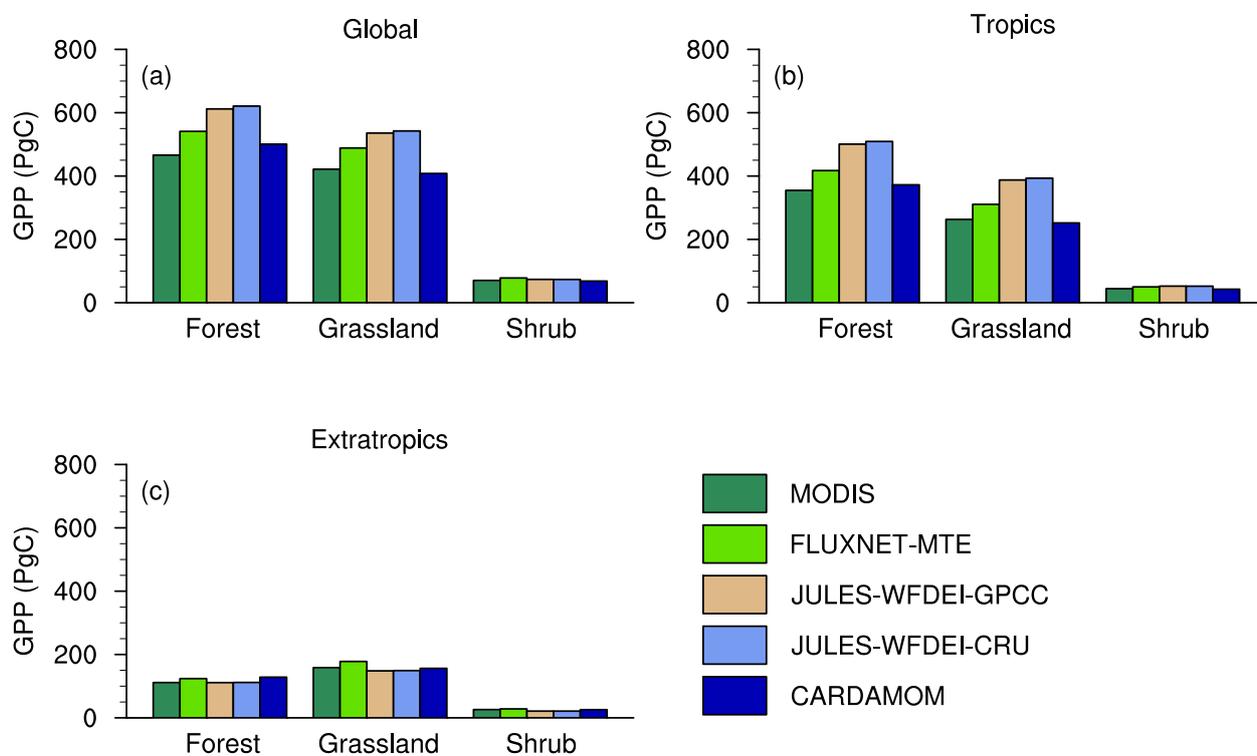
**Figure 1.** Map showing the regions specified in Table 2.



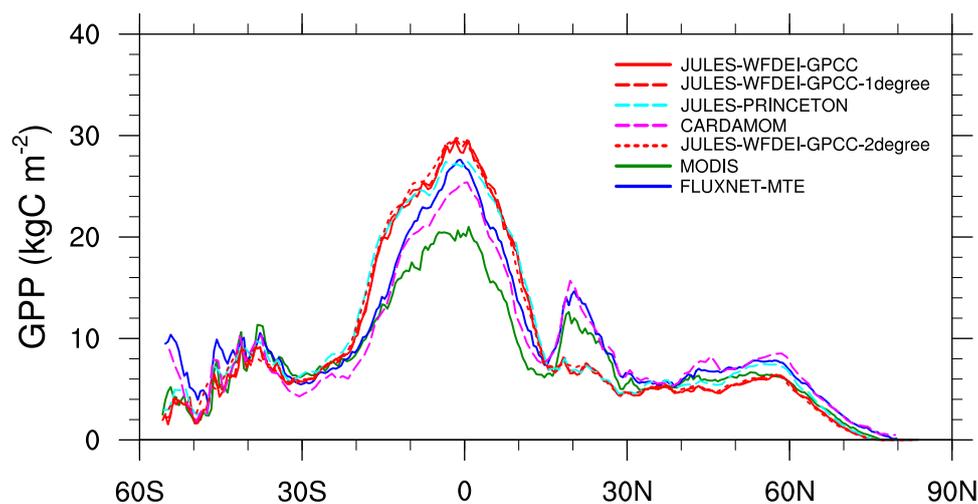
**Figure 2.** Comparison of JULES, observation-based (FLUXNET-MTE and MODIS) and CARDAMOM (Table 1) GPP fluxes for the 2001–2010 period at global scales. **(a)** shows the global area-weighted average of the mean monthly GPP, **(b)** shows the coefficient of variation (CV) expressed as percentages of the mean monthly GPP and **(c)** shows the monthly anomalies (global area-weighted mean) expressed as percentages of the mean monthly GPP (global area-weighted mean) for each month.



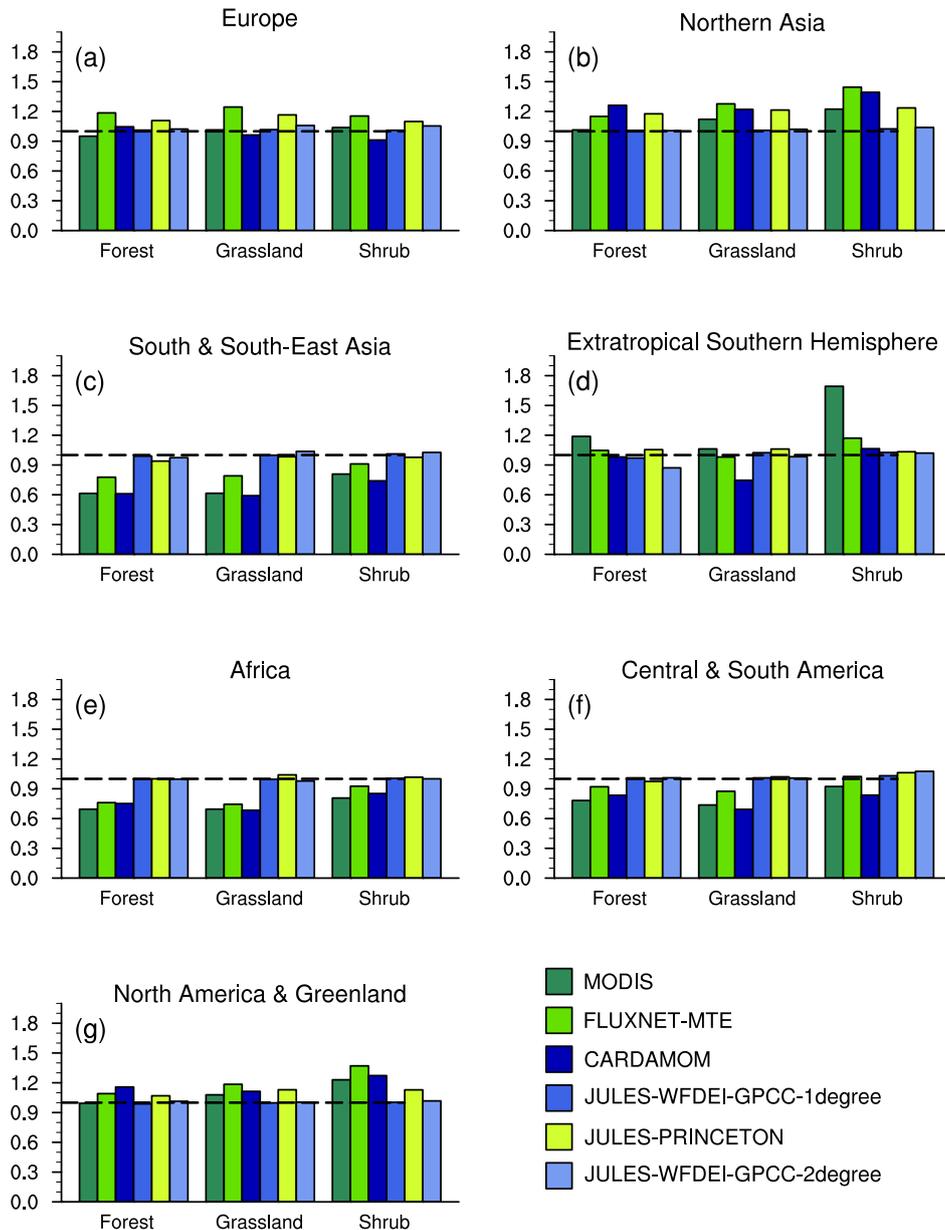
**Figure 3.** Total annual and zonal mean model simulated (JULES-WFDEI-GPCC), observed (FLUXNET-MTE and MODIS) and CARDAMOM GPP fluxes for the 2001–2010 period at the global scale. JULES, FLUXNET-MTE and MODIS GPP are at  $0.5^\circ \times 0.5^\circ$  spatial resolution and CARDAMOM is at  $1^\circ \times 1^\circ$  resolution. (a), (b), (c) and (d) show the total annual GPP of JULES-WFDEI-GPCC, FLUXNET-MTE, MODIS and CARDAMOM GPP, respectively. At the top right of each map subplot, the average global annual GPP for 2001–2010 is displayed. (e) shows the zonal mean of the total annual JULES-WFDEI-GPCC, FLUXNET-MTE, MODIS and CARDAMOM GPP, respectively. Included in each map subplot are contour lines for the tropical regions.



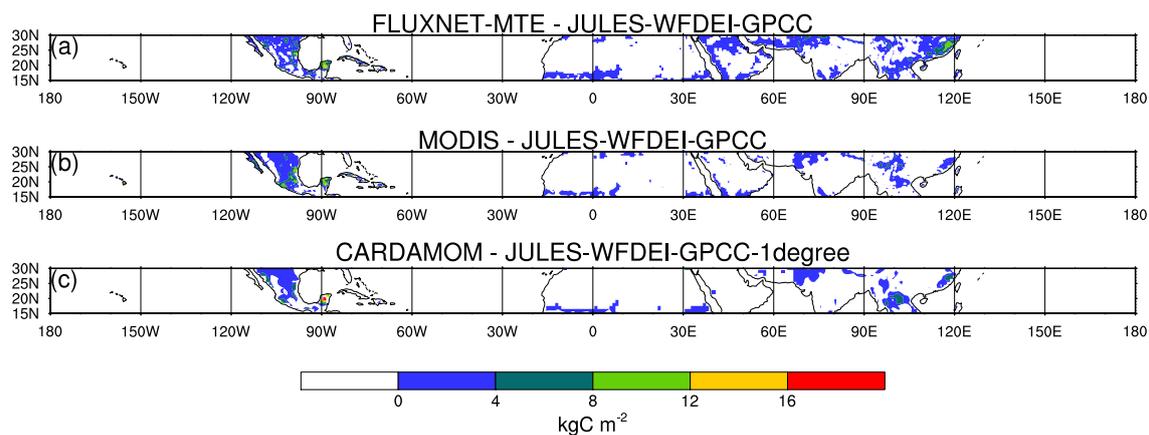
**Figure 4.** Total (summed over 10 years) model simulated (JULES-WFDEI-GPCC, JULES-WFDEI-CRU and CARDAMOM), observation-based (FLUXNET-MTE and MODIS) GPP fluxes for the 2001–2010 period at global and regional scales (tropics and extratropics) for 3 biome types (Forest, Grassland and Shrub). (a) shows the global total annual GPP, (b) for the tropics (30°S–30°N) and (c) for the extratropics (30°N–90°N and 30°S–90°S) for forests, grasslands and shrubs.



**Figure 5.** Zonal mean of total annual model simulated (JULES-WFDEI-GPCC, JULES-WFDEI-GPCC-1degree, JULES-PRINCETON, CARDAMOM and JULES-WFDEI-GPCC-2degree) and observed (FLUXNET-MTE and MODIS) GPP fluxes for 2001–2010. JULES-WFDEI-GPCC, FLUXNET-MTE and MODIS are at  $0.5^\circ \times 0.5^\circ$  spatial resolution.



**Figure 6.** Total annual model simulated (JULES-WFDEI-GPCC-1degree, JULES-PRINCETON, CARDAMOM and JULES-WFDEI-GPCC-2degree) and observed (FLUXNET-MTE and MODIS) GPP fluxes for the 2001–2010 period normalised by model simulated (JULES-WFDEI-GPCC) GPP for various regions (Table 2) for 3 biome types (Forest, Grassland and Shrub). (a) shows normalised GPP for Europe, (b) for Northern Asia, (c) for South & South-Asia, (d) for extratropical Southern Hemisphere, (e) for Africa, (f) for Central & South America and (g) for North America & Greenland. The dotted line at  $y=1$  represents where the model and observations match.



**Figure 7.** Difference in total annual GPP between JULES-WFDEI-GPCC and the observation-based (FLUXNET-MTE and MODIS) and CARDAMOM estimates of GPP for the 2001–2010 period at latitudes 15°N–30°N. (a) shows the difference between FLUXNET-MTE and JULES, (b) between MODIS and JULES and (c) between CARDAMOM and JULES. A positive change in GPP means the observation-based estimates (FLUXNET-MTE and MODIS) or CARDAMOM estimate are higher than the model.



**Table 1.** Types of global scale model simulations performed.

Model simulations	Meteorological forcing	Spatial resolution	Grid dimensions <sup>a</sup>
JULES-WFDEI-GPCC	WFDEI-GPCC	$0.5^\circ \times 0.5^\circ$	$720 \times 360$
JULES-WFDEI-CRU	WFDEI-CRU	$0.5^\circ \times 0.5^\circ$	$720 \times 360$
JULES-WFDEI-GPCC-1degree	WFDEI-GPCC	$1^\circ \times 1^\circ$	$360 \times 180$
JULES-PRINCETON	PRINCETON	$1^\circ \times 1^\circ$	$360 \times 180$
JULES-WFDEI-GPCC-2degree	WFDEI-GPCC	$2^\circ \times 2^\circ$	$180 \times 90$

<sup>a</sup> Grid dimensions are given as the number of grid boxes in the longitudinal direction by the number of grid boxes in the latitudinal direction.



**Table 2.** List of regions used. Only land grid points are used in the analysis.

Name	Latitude (°)	Longitude (°)
Europe	30N–90N	15W–45E
Northern Asia	30N–90N	45E–180E
South & South-East Asia	30S–30N	60E–150E
Extratropical Southern Hemisphere	60S–30S	120W–180E
Africa	30S–30N	30W–60E
Central & Southern America	30S–30N	120W–30W
North America & Greenland	30N–90N	180W–15W