Effect of climate dataset selection on simulations of terrestrial GPP: Highest uncertainty for tropical regions

Zhendong Wu[1,2]*, Niklas Boke-Olén[1], Rasmus Fensholt[2], Jonas Ardö[1], Lars Eklundh[1], Veiko Lehsten[1,3]

1 Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden, 2 Department of Geosciences and Natural Resource Management, University of Copenhagen, Copenhagen, Denmark, 3 Swiss Federal Institute for Forest, Snow and Landscape research (WSL), Birmensdorf, Switzerland

* zhendong.wu@nateko.lu.se

Abstract

Biogeochemical models use meteorological forcing data derived with different approaches (e.g. based on interpolation or reanalysis of observation data or a hybrid hereof) to simulate ecosystem processes such as gross primary productivity (GPP). This study assesses the impact of different widely used climate datasets on simulated gross primary productivity and evaluates the suitability of them for reproducing the global and regional carbon cycle as mapped from independent GPP data. We simulate GPP with the biogeochemical model LPJ-GUESS using six historical climate datasets (CRU, CRUNCEP, ECMWF, NCEP, PRINCETON, and WFDEI). The simulated GPP is evaluated using an observation-based GPP product derived from eddy covariance measurements in combination with remotely sensed data. Our results show that all datasets tested produce relatively similar GPP simulations at a global scale, corresponding fairly well to the observation-based data with a difference between simulations and observations ranging from -50 to 60 g m^{-2} yr^{-1}. However, all simulations also show a strong underestimation of GPP (ranging from -533 to -870 g m^{-2} yr^{-1}) and low temporal agreement (r < 0.4) with observations over tropical areas. As the shortwave radiation for tropical areas was found to have the highest uncertainty in the analyzed historical climate datasets, we test whether simulation results could be improved by a correction of the tested shortwave radiation for tropical areas using a new radiation product from the International Satellite Cloud Climatology Project (ISCCP). A large improvement (up to 48%) in simulated GPP magnitude was observed with bias corrected shortwave radiation, as well as an increase in spatio-temporal agreement between the simulated GPP and observation-based GPP. This study conducts a spatial inter-comparison and quantification of the performances of climate datasets and can thereby facilitate the selection of climate forcing data over any given study area for modelling purposes.

Introduction

Biogeochemical models are widely used to refine and upscale field measurement of spatiotemporal carbon exchange and great advances have been made in developing these models in the last decade (e.g. [1–4]). Furthermore, biogeochemical models are used to predict future carbon budgets under different scenarios providing descriptions of future potential biogeochemical conditions essential to assess socioeconomic, technological and environmental conditions, emissions of greenhouse gases and aerosols, as well as climate [5, 6]. These models are usually driven by climate data and simulate the spatio-temporal vegetation dynamics as well as the carbon fluxes and water flows through the ecosystem [7–9]. However, the choice of historical climate dataset input can cause considerable uncertainty in estimated Gross Primary Production (GPP, the total amount of carbon captured by vegetation via photosynthesis) with outputs fluctuating by 9% to 20% [10–12]. The choice of the climate dataset also has a pronounced impact on the spatial patterns of simulated GPP [11, 13]. Therefore, the selection of historical climate datasets plays a crucial role in both exploring and quantifying the ecosystem response to climate through ecosystem models.

Uncertainty among different historical climate datasets exist at present, which mainly differ in the source and the processing of the raw data. Such climate grids are derived either from quasi-point based measurements and subsequent spatial interpolation, model-based reanalysis, or generated as an observational-reanalysis hybrid. Measurement-based datasets, e.g. Climatic Research Unit (CRU; [14]), are produced by statistical interpolation of climate station records, e.g. by using the Climate Anomaly Method [15]. Reanalysis is a different approach that uses a combination of meteorological forecast model output and assimilated observations. Unlike the observational based datasets, which are based on statistical principles, reanalysis datasets are built on physical principles describing the variable in question [16], by combining climate model output with a large amount of different observational data, such as land cover, trace gases, aerosols, solar variations and wind speed. As a third type, observational-reanalysis hybrid datasets combine observations and reanalysis data [17, 18].

This study is motivated by two factors: Firstly, as there is no general agreement about which historical climate data set is most suitable for driving biogeochemical models, several of the currently available historical climate datasets are widely used for contemporary research on estimations of GPP [2, 4, 19–21], yet very few studies (e.g. [12,13]) have investigated the difference in reproducing the carbon cycle associated with the use of climate datasets. The suitability of contemporary historical climate datasets for accurately estimating GPP at global and regional scales is therefore currently not well known. Secondly, users of biogeochemical models normally rely on the climate dataset for which the model was calibrated to reproduce the carbon cycle with the least uncertainty for a particular region. However, climate datasets might vary in quality in a spatially explicit way governed by the processing algorithm and underlying density of available calibration points. There are also incidences in which the user can not choose the dataset for which the model was calibrated, e.g. in a model comparison study where different models need to be driven by similar input data, or if the calibration dataset has a lower temporal resolution than what is required by a specific task. Therefore, this study fills a current research gap by evaluating the six most commonly used climate datasets (CRU, CRUNCEP, ECMWF, NCEP, PRINCETON, WFDEI; See Methods) and their relative performance of estimating terrestrial GPP within a spatially explicit biogeochemical model to highlight the associated uncertainty. Such quantification is expected to facilitate the selection of relevant climate forcing data when performing GPP modelling over any given study area.

Here we focus on terrestrial GPP, a fundamental driver of plant biochemical processes and an important component of the global carbon cycle [22]. In biogeochemical models, GPP represents the origin of carbon within the system, which controls many other processes (e.g.

carbon allocation, plant allometry and tissue turnover) in the models [2, 4, 19–21]. GPP is mainly influenced by climate forcing (e.g. temperature, water, light, and atmospheric CO_2 concentration), and also influenced by nutrient availability and disturbances (e.g., storms, harvesting, and insect attacks). We use the biogeochemical model Lund-Potsdam-Jena General Ecosystem Simulator (LPJ-GUESS [3, 21]) as a representative model to simulate GPP and compare results to an independent observation-based GPP product, (even though climate data is also used to generate the observation-based GPP product; see discussion for more details). The observation based GPP product is derived from a global network of eddy covariance measurements in combination with remote sensing data and is used as a benchmark to evaluate the performances of the climate datasets. We analyze the differences in magnitude and spatio-temporal pattern of GPP globally and over five vegetated land cover classes to assess their relative performance of reproducing the carbon cycle during the period 1982–2010.

Methods

Biogeochemical model (LPJ-GUESS)

LPJ-GUESS is a process-based biogeochemical model, designed for both regional and global studies [21]. It requires time series data of climate forcing (i.e. air temperature, precipitation and shortwave radiation) and atmospheric carbon dioxide concentrations as input. It explicitly represents vegetation cover (by indicating the occurrence of Plant Functional Types, PFTs), age cohorts, gap dynamics and biogeochemical cycles. Vegetation physiological processes such as photosynthesis, canopy conductance, phenology, and carbon allocation are incorporated in the model. LPJ-GUESS uses a detailed individual-based representation of forest stand structure and dynamics for PFTs co-occurring in a number of patches or local stands, representative for the landscape of a grid cell. Each PFT is characterized by properties such as growth form, leaf phenology, life history and bioclimatic limits, which govern their performance and competitive interactions under the forcing conditions and realized ecosystem state of a particular grid cell [20, 23]. In total 11 PFTs are used within this study and their prescribed parameters can be found in Smith *et al.* [3]. We employ LPJ-GUESS version 3.0 [3] which uses nitrogen dynamic based on the CENTURY model [24, 25]. All simulations are initialized with a 500 years spin-up, which comprises an internal 40000 years spin-up mechanism for soils, to equilibrate soil and vegetation pools, by recycling de-trended 1979–2010 climate forcing fields and applying constant CO_2 concentration and nitrogen deposition from the first year (1979). Subsequently transient GPP is simulated with time evolving CO_2 concentrations from Keeling and Whorf [26], nitrogen deposition from Lamarque *et al.* [27] and climate forcing. The managed land use fraction is obtained from Hurtt *et al.* [28].

Historical climate datasets

We force LPJ-GUESS with six different historical climate datasets (Table 1) to simulate global terrestrial GPP. The datasets differ in their spatial and temporal resolution, available time period, and how they are derived. They are derived from quasi-point based measurements (CRU and CRUNCEP), model-based reanalysis (NCEP and ECMWF), or hybrid datasets combining both observation and reanalysis data (WFDEI and PRINCETON). To enable a direct comparison between simulations, the datasets are rescaled to a common spatial (0.5 decimal degree) and temporal scale (monthly observations, since CRU is provided only as monthly data). To allow this, we use bilinear interpolation to convert NCEP to 0.5 degrees, and temporally convert CRUNCEP, ECMWF, NCEP, PRINCTON and WFDEI from daily to monthly time scales. These monthly datasets are subsequently interpolated to daily values uniformly within LPJ-GUESS [21]. Since LPJ-GUESS treats all dataset in the same way, it offsets

Table 1. Main datasets used, type, spatial resolution and time period.

Dataset	Type	Spatial resolution	Time period	Reference
CRU TS 3.21	Climate	0.5 degree	1901–2012	[14]
CRUNCEP v5	Climate	0.5 degree	1901–2013	[30]
ECMWF/ERA Interim	Climate	0.5 degree	1979–2014	[31]
NCEP-DOE II	Climate	2.5 degree	1979–2014	[32]
Princeton_V2	Climate	0.5 degree	1901–2012	[17]
WFDEI_GPCC	Climate	0.5 degree	1979–2010	[18]
ISCCP	Radiation	0.5 degree	1984–2000	[33]
JUNG11	GPP	0.5 degree	1982–2011	[34]

https://doi.org/10.1371/journal.pone.0199383.t001

at least part of interpolation induced bias. We use the common time period 1979–2010 and the climate variables precipitation, shortwave radiation, and air temperature for all datasets. For CRU, the cloud cover is converted to shortwave radiation within the biogeochemical model using the method by Harris *et al.* [29]. All data are available on the DataGURU server (https://dataguru.lu.se/).

Observation-based GPP product

We evaluate the simulated GPP with a benchmark GPP product derived from eddy covariance measurements from Jung *et al.* [34] (herein after, JUNG11). JUNG11 is derived from long-term and high-quality measurements of carbon dioxide, water, and energy fluxes from the Flux Network (FLUXNET). These *in situ* measurements are very sparse at the global scale, and need to be extrapolated in space, in order to be applicable for global scale studies. Jung *et al.* [34] used a semi-empirical model (Model Tree Ensembles; MTE), to upscale measurements from local to global scales using remotely sensed fraction of Absorbed Photosynthetically Active Radiation (fAPAR), gridded climate, and the Synergetic land cover product (SYN-MAP). The long-term mean climatic information used in JUNG11 is derived from CRU data [35] as well as other climate datasets, e.g. global grids of monthly precipitation from GPCC [36] and the ECMWF ERA interim reanalysis product of Simmons *et al.* [37]. In this study, the observation-based JUNG11 dataset is assumed to represent "true" information of GPP, though we are well aware of the uncertainties related to this product, e.g. the uncertainties originating from flux measurements and upscaling station-based fluxes to global scale [34].

Comparison of GPP estimates with the different climate datasets

Two of the most commonly used metrics to compare model estimates with observations, are the Pearson correlation coefficient (r) and root mean square deviation (RMSD). Despite their popularity, both metrics have disadvantages, as r only measures the strength of relationship between two data series, but does not indicate if the data series have similar magnitude. RMSD, on the other hand, assesses if the absolute values of two series match, but does not indicate the agreement of pattern of the data series. Moreover, RMSD is dimensional, which hampers inter-comparability between analysis outputs. To consider both the strength of the relationship and similarity in magnitude, Willmott [38] proposed an index of agreement (IoA, *d*) for evaluating model prediction (*P*) against measured observations (*O*), as follows:

$$d = 1 - \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}(|P_i - \overline{O}| + |O_i - \overline{O}|)^2} \tag{1}$$

The upper limit of IoA (*d*) is one which indicates a perfect match, while the lower limit is zero which indicates complete disagreement. The metric describes the relative co-variability of *P* and *O* related to the observed mean ($\overline{O}$).

We use Willmott's IoA to quantify the match between simulated GPP and JUNG11. If two datasets differ by only 5% we define their result as equal to allow for some small statistical differences. Furthermore, we use the correlation coefficient (r) and annual means as an evaluation measure to assess the IoA result and link it to temporal patterns or magnitude differences. The comparisons are conducted globally and for five vegetated land cover classes derived from Ahlström *et al.* [39] during 1982–2010 (S1 Fig).

Radiation correction

Simulated GPP has its largest deviation among the datasets for the tropical region (according to initial calculations; S2 Fig). A previous study [12] also revealed that climate dataset induced uncertainty in GPP estimates simulated by LPJ-GUESS was mainly caused by the uncertainty in shortwave radiation over tropical regions. Therefore, we test whether bias correcting the shortwave radiation variable of the tested climate datasets using the International Satellite Cloud Climatology Project (ISCCP) radiation data results in an improvement of the simulated GPP. The ISCCP radiation product is derived from an advanced radiative transfer model (NASA Goddard Institute for Space Studies) by using improved cloud climatology and ancillary data sets [33, 40]. ISCCP has been used as the reference radiation in previous studies, e.g. [41, 42].

The bias correction of shortwave radiation is done only for tropical regions, where simulated GPP is particularly sensitive to shortwave radiation [12], the largest deviation in simulated GPP is shown (S2 Fig) and the largest difference in shortwave radiation among the datasets is present. The differences of the monthly mean shortwave radiation between the tested climate datasets and the ISCCP radiation during the common 17 years (1984–2000) is used to correct the original monthly data from the tested climate datasets (R_t^{orig}) during 1982–2010 using Eq 2.

$$R_t^{corr} = R_t^{orig} + \left(\overline{R^{ref}} - \overline{R^{orig}}\right), \tag{2}$$

where R_t^{corr} is the bias corrected shortwave radiation for month *t*. $\overline{R^{orig}}$ and $\overline{R^{ref}}$ are the monthly mean of the tested climate datasets and the ISCCP radiation during 1984–2000, respectively. This bias correction adjusts for biases in annual averages and seasonal distribution, while preserving the inter-annual variability.

Results

The agreement between the simulated GPP and the GPP provided by JUNG11 is compared for each grid cell at a 0.5 degree scale using IoA (Fig 1). Our result shows that simulations of GPP using CRU climate data (CRU GPP) have the highest spatial agreement with the reference dataset (31% of global vegetated grid cells where CRU GPP produce the highest IoA). Areas of best agreement are mainly located in the Northern boreal forest but large clusters are also observed in parts of Europe and the United States (Fig 1A). We further found that for a majority of the area (40%, mainly occupied by tropical and dry area), one single dataset was identified with an agreement at least 5% higher (as measured by IoA) than the other datasets (marked as green in Fig 1B). We also found a low IoA for the tropical forest region (Fig 1C), namely tropical Asia, central Africa and tropical South America as well as for desert areas in Africa and Australia. This is in accordance with the simulated GPP showing an expected low
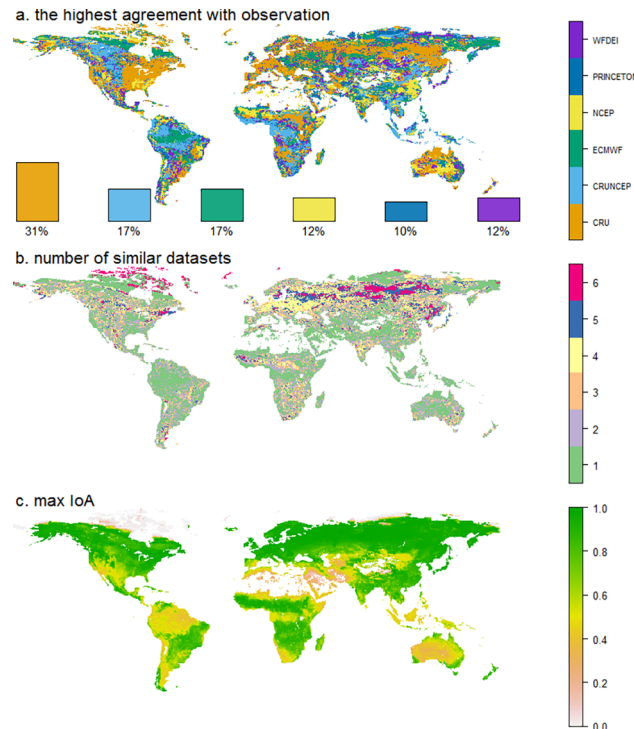
Fig 1. Global maps of climate dataset performance. Panel a) indication of which climate dataset is producing the highest Index of agreement (IoA; calculated at monthly scale) to JUNG11 (1982–2010). The bars show a global total fraction of vegetated grid cells for which the climate dataset is giving the highest IoA. Panel b) shows how many datasets producing GPP simulations with a similar agreement (within 5%) as the one identified in (a). Panel c) displays the maximum IoA between simulated GPP and JUNG11 for each grid cell.

https://doi.org/10.1371/journal.pone.0199383.g001

mean IoA (<0.5) for Tropical Forests (TF) (Fig 2A) across all datasets with a large bias (Fig 2B) as well as poor temporal correlation (Fig 2C), among which CRUNCEP performs slightly better (r = 0.53). However, at a global scale, the simulated GPP magnitude is relatively close to JUNG11 estimates (Fig 2B). On average simulated GPP is only 6 g m^{-2} yr^{-1} higher than JUNG11, indicating a compensation of regional discrepancies according to overestimation in non-TF and underestimation in TF.

The comparison of climate data inputs (Fig 3) shows that the zonal mean of the annual temperature is similar among the six climate datasets (Fig 3A) and the precipitation datasets also agree relatively well (Fig 3B) except around the equator and in latitudes below 30-degree South which can be partly attributed to the small number of grid cells in that region. The highest variability is found for the shortwave radiation data (Fig 3C) where the largest discrepancies are found in low latitudes (20°S-20°N), with the CRU shortwave radiation standing out with an average ~14% lower value compared to the mean of the other datasets.

To evaluate the influence of the shortwave radiation on the simulated GPP for tropical forest (TF), we bias corrected the shortwave radiation datasets using ISCCP data [33]. The corrected shortwave radiation in tropical forests increased for all climate datasets as compared to the original shortwave radiation. The average increase is lowest for CRUNCEP [30] with 4.9 W m^{-2} and highest for CRU [14] with 57.4 W m^{-2} (S3E Fig). By using the bias corrected shortwave radiation (bar with black outline, Fig 4) over the tropical forest, the average annual GPP shows a 7.5% overall increment while CRU increases with 16.8% and ECMWF shows an increase of 11.7%, and GPP is on average 23.0% closer to the JUNG11 value for all datasets (Fig 4B). This increase in agreement is especially pronounced for the simulation using

Fig 2. Comparison of monthly IoA, annual mean GPP and monthly temporal correlation during 1982–2010 as estimated by LPJ-GUESS forced by six climate datasets versus the observation-based GPP product JUNG11. Panel (a) shows the IoA, panel (b) shows the average difference and the last panel (c) shows the temporal correlation coefficient between simulated GPP and observations for each land cover class: global; semi-arid ecosystems (SS); tundra and arctic shrub land (TS); grasslands and land under agriculture (GC); tropical forest (TF); extra-tropical forest (ExTF) includes boreal and temperate. The map of land cover classes can be seen in S1 Fig. The spatial distribution of GPP magnitude can be found in S2 Fig.

ECMWF climate data [31], being 48.1% closer to the JUNG11 value after bias correction. There is no significant difference (p>0.05, using one-way ANOVA test) between CRUNCEP simulations with and without bias correction, further confirming that CRUNCEP shortwave radiation has the smallest deviation from the ISCCP data in tropical forest, which could partly explain why CRUNCEP has the relatively higher agreement with JUNG11 over tropical forest in Fig 2A.
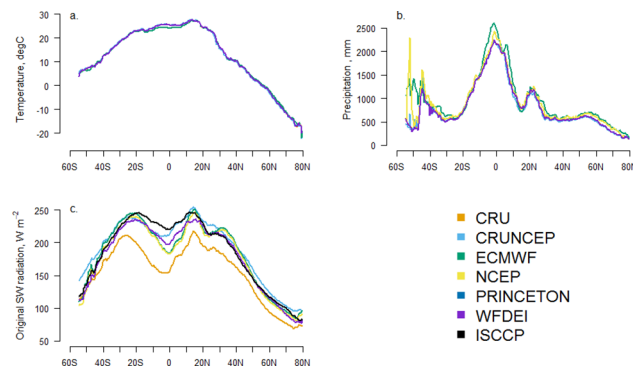

Fig 3. Comparisons of the climatological zonal mean of annual average (1982–2010) of three climate variables among the six climate datasets, i.e. CRU [14], CRUNCEP [30], ECMWF [31], NCEP [32], PRINCETON [17] and WFDEI [18]. The variables are temperature (panel a), precipitation (panel b) and shortwave radiation (panel c). The black line in panel (c) shows the zonal mean of annual average (1984–2000) of ISCCP radiation [33]. The comparisons are conducted for terrestrial areas only. The spatial distribution of each climate variable can be found in S4–S6 Figs.
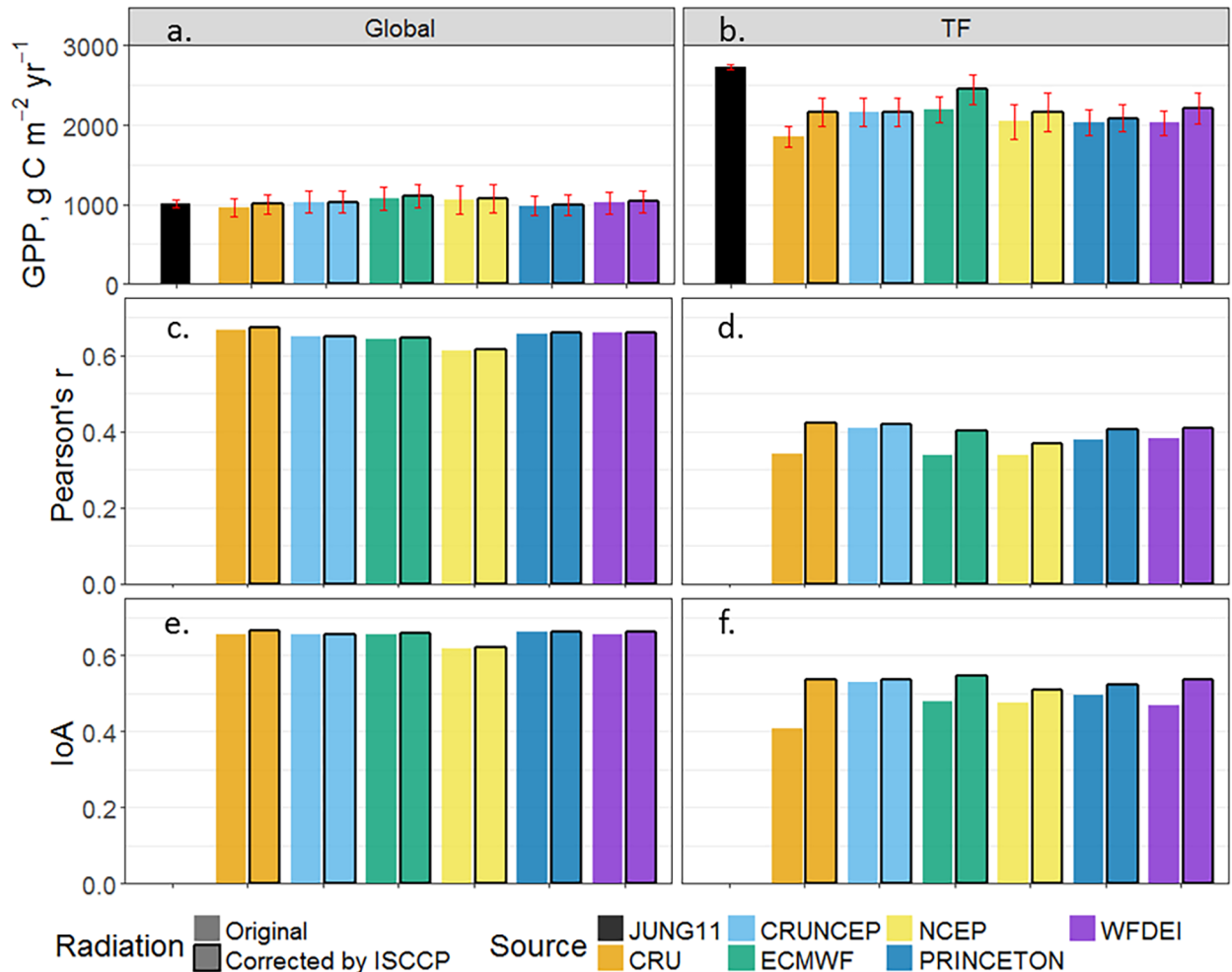
Fig 4. Comparison of annual mean GPP, monthly temporal correlation and monthly IoA during 1982–2010 estimated by LPJ-GUESS before and after tropical forest radiation correction. Panels (a-b) show annual mean GPP, panels (c-d) show temporal correlation and panels (e-f) show the IoA. Bars with a black outline represent simulations based on shortwave radiation corrected by ISCCP data. The extent of tropical forest (TF) is shown in S1 Fig. The red error bars show the inter-annual variability.

The overall effect of correcting shortwave radiation over the tropical forest on simulated global annual mean GPP (Fig 4A) is only 1.2% compared to simulations based on uncorrected shortwave radiation data. The effect of the radiation correction on the temporal correlation (0.6%) and IoA (0.7%) is also negligible on the global scale (Fig 4). The climate induced spread of simulated GPP among climate datasets tested at global scale was reduced from 11.0% to 10.8% by correcting shortwave radiation over the tropical forest.

Discussion

This study evaluates six climate datasets and their influence on gross primary productivity (GPP) simulated by a biogeochemical model (LPJ-GUESS). Given that GPP is the main driver for a number of vegetation based processes our results can also help to improve the estimation of a variety of other state variables (e.g. net primary productivity). LPJ-GUESS is a well-established biogeochemical model that has been evaluated and applied in a wide range of studies and shows relatively similar behavior and predictive skills compared to other biogeochemical

models [43–45]. This is especially true for GPP, given that most biogeochemical models (e.g. HYLAND, LPJ-DGVM, OCN, ORCHIDEE, SDGVM and TRIFFID) use the same photosynthesis model [46] at their core [9, 39, 47]. LPJ-GUESS may thus be considered a generic representative for biogeochemical models as a group and very likely reproduces spatial and temporal characteristics of primary productivity.

The datasets investigated are all widely used in studies focusing on modeling the carbon cycle and the results show that the differences are most pronounced for shortwave radiation. CRU shortwave radiation is calculated from cloud cover, which is derived from observations of sun hours, by using the method of Harris *et al.* [29]. CRUNCEP shortwave radiation on the other hand is rescaled from the NCEP-NCAR [48] reanalysis data by using the MTCLIM model [49], which reduces the magnitude of NCEP-NCAR shortwave radiation to better match observed radiation at FLUXNET sites [30]. The reanalysis shortwave radiation from ECMWF and NCEP (here we use DOE II which differs from NCEP-NCAR) are produced by different radiative transfer schemes from Mlawer *et al.* [50] and Chou [51], respectively. These schemes describe how solar irradiance is attenuated by the absorption and scattering (due to e.g. water vapor, oxygen, trace gases, clouds, and aerosols) when passing through the atmosphere before reaching the land surface. PRINCETON shortwave radiation is based on interpolating the NCEP-NCAR reanalysis product and downscaling to 0.5 degree prior to bias correction using CRU data [17]. WFDEI shortwave radiation is derived from the ERA-40 reanalysis product [52] and the dataset is adjusted by using CRU cloud cover [53]. Given the manifold methodological differences, it is a challenge to determine whether all datasets are equally reliable or if any of them is better suited for a certain study region or purpose than others.

Overall, CRU driven GPP results in the best agreement with JUNG11 for the largest area compared to the other climate datasets, which may be due to the fact that the JUNG11 has been generated by incorporating CRU data to some extent [34]. However, still in almost 70% of the vegetated area JUNG11 agrees better with one of the other climate datasets used as input for simulating GPP. We also used the observation-based MODIS GPP product [54] in our analyses. Even that the MODIS GPP algorithm includes the NCEP dataset (one of tested datasets) as an input of daily meteorological data, the results agreed that CRU is a better climate forcing in more grid cells than the other datasets tested (S7 Fig). Considering that the long-term observation and climatic information used in JUNG11 is not entirely from CRU, we decided to use JUNG11 as the benchmark of this study. Our study shows that the specific choice of the climate dataset to be used for driving the biogeochemical model (out of the six historical datasets investigated) is associated with smaller spread in simulated GPP at the global scale than the spread at the regional scale (Fig 2), which indicates that the choice of the climate dataset for estimating global GPP is less critical as when estimating GPP at the regional scale. The largest disagreement of GPP between LPJ-GUESS simulations and JUNG11, is found in the tropical region. This pattern is consistent with findings that the tropical region has the largest differences in GPP estimates between process-based models and data-driven methods [55–57]. We also found the largest disagreement between simulated GPP in the tropical region, which is attributed to the large bias of shortwave radiation among investigated climate datasets and the high sensitivity of GPP to shortwave radiation over the tropics [12]. Wu *et al.* [12] also showed that differences in shortwave radiation caused large differences in simulated GPP over tropical regions when using LPJ-GUESS, which is likely to be similar for other biogeochemical models [10]. The bias of shortwave radiation in tropical areas has been attributed to the sparse meteorological station network [58] and to the high uncertainties in radiation transfer, cloud cover and cloud morphology when producing the climate datasets [11, 59].

We also show that a bias correction of shortwave radiation data (using the ISCCP radiation data) in the climate datasets causes the simulated GPP to markedly increase in the tropical

region (e.g. CRU and ECMWF simulations), reducing the gap in simulated GPP compared to JUNG11. This again suggests that shortwave radiation products currently available for tropical regions remain highly uncertain. In order to accurately simulate GPP in tropical regions (which are known to be primarily constrained by incoming solar radiation) we suggest improving shortwave radiation of the tested datasets, e.g. by bias correcting with advanced radiation data from ISCCP. ISCCP reduced cloud effects on radiation by using an advanced radiative transfer model [33, 40], which makes ISCCP radiation data more reliable in cloud-prone tropical forest areas than the radiation from the tested climate datasets (except CRUN-CEP). Following the bias correction, we found no significant change for the CRUNCEP simulation, which suggested that the shortwave radiation from CRUNCEP has equally high quality as ISCCP. The high quality of the shortwave radiation data from CRUNCEP in the tropic is likely to be one of the reasons for more grid cells of highest IoA being derived from the CRUN-CEP stimulation in Fig 1A. Furthermore, we found that correcting only for shortwave radiation is not enough to produce an exact match with observation-based estimates as there is still a substantial gap between model simulations and JUNG11. The ECMWF dataset, character-ized by the highest precipitation, also showed the highest agreement with JUNG11 after short-wave radiation correction, which implies that not only the radiation but also the precipitation over tropical areas might be underestimated in the climate datasets tested. Previous studies [10, 12] also found that GPP was sensitive to precipitation in tropical areas. Therefore, if aim-ing at producing a set of climate variables to minimize the discrepancy between modelled and observed GPP, we recommend also to improve the precipitation variable of tested climate datasets (e.g. CRUNCEP which has high quality radiation and temperature data) e.g. by bias correction using high quality precipitation data (e.g. TRMM [60]) in tropical region.

Although correcting the shortwave radiation over the tropical forest reduced the climate induced spread of simulated GPP among climate datasets tested at global scale from 11.0% to 10.8%, which is within the range of 9%-20% [10–12], the aim of the bias correction was not to narrow the climate induced spread. We would expect that if correcting all of the three climate variables there will be no climate induced spread among climate datasets tested. Bias correc-tion is one way that could help improving the climate variable of a climate dataset in a certain study area, by using ISCCP, TRMM or other available high quality data. However, the correc-tion of a given climate variable within a climate dataset should be done with caution, as improving a single variable from a climate dataset may introduce an imbalance in relation to other co-varying climate variables of that dataset. Therefore, we consider it preferably to first select a suitable climate dataset for a study area and then, if deemed necessary, a given variable of this dataset can additionally be bias-corrected.

In order to avoid over-interpretation of model-data mismatches, it is mandatory to also consider the limitations of the reference data. JUNG11 GPP used in this study was assumed to represent the "true" GPP but inevitably also includes systematic and random errors and uncer-tainties. For instance, uncertainties of flux measurements derived from discriminating low and well mixed fluxes [61], estimation of missing values [62], and flux partitioning (e.g. parti-tion the observed net ecosystem exchange (NEE) in to GPP and ecosystem respiration) [63, 64]. These uncertainties, furthermore, propagate when extrapolating to the globe by the MTE approach [34]. One additional complication arises from the possibility that JUNG11 also per-forms poorly over tropical areas and that disentangling uncertainties within the GPP simulated by LPJ-GUESS and JUNG11 might be impossible.

One additional limitation of this study is related to the evaluation method. IoA is used as the main metric for the evaluation since it combines patterns like the Pearson correlation coef-ficient and information on the magnitude of deviations. However, it is known to be sensitive to extreme values due to the squared differences which potentially over-weighs the influence

of the differences between model prediction and observation [65]. Hence, we have complemented this statistical measure with calculations of the average difference and correlation coefficient.

Conclusion

This study evaluates the performance of the six most commonly used climate datasets (CRU, CRUNCEP, ECMWF, NCEP, PRINCETON, WFDEI) in estimating terrestrial GPP within a spatially explicit biogeochemical model by using independent observation-based GPP data. Our study highlights the need to improve the incoming shortwave radiation estimates from most of the climate datasets tested (except CRUNCEP) in tropical areas in order to improve GPP estimates over tropical regions. Our results also allow the assessment of the suitability of climate datasets with respect to a given research purpose and study area, e.g. the CRUNCEP dataset works better in tropical regions for simulating GPP (values being in agreement with observation-based GPP), while the choice of the climate dataset for simulating GPP in Europe is less critical.

Supporting information

S1 Fig. Map of land cover classes. The source of the data derived from Ahlström et al. [39] and Wu et al. [12]. The percentage values at the bottom of the map show the fraction of each land cover class in relation to the global terrestrial area (excluding Greenland).
(TIF)

S2 Fig. Comparison of annual mean GPP during 1982–2010 from model simulations by using different climate datasets and observation-based estimate (JUNG11). a. global GPP linear trends. b. GPP zonal means. c-h maps of spatial difference of annual mean GPP between simulations forced with different climate datasets and observations (g C /m^{-2}).
(TIF)

S3 Fig. Annual mean shortwave radiation during 1982–2010 globally and stratified by land cover classes. Bars with a black outline represent the simulations based on shortwave radiation is corrected by ISCCP data.
(TIF)

S4 Fig. Comparison of annual temperature from the climate datasets tested. a. global annual trends, b. zonal means, c-j. spatial distribution of mean annual temperature.
(TIF)

S5 Fig. Comparison of annual total precipitation from the climate datasets tested. a. global annual trends, b. zonal means, c-j. spatial distribution of mean annual precipitation.
(TIF)

S6 Fig. Comparison of annual shortwave radiation from the climate datasets tested. a. global annual trends, b. zonal means, c-j. spatial distribution of mean annual shortwave radiation.
(TIF)

S7 Fig. Global maps of climate dataset performance. Panel a, c, and e show the results when using MODIS GPP (2000–2010) as the benchmark, and panel b, d and f show the results when using JUNG11 GPP (1982–2010) as the benchmark. For the description of the figure is referred to Fig 1 in the main text.
(TIF)

Acknowledgments

Author Contributions

Conceptualization: Zhendong Wu, Veiko Lehsten.

Data curation: Zhendong Wu.

Formal analysis: Zhendong Wu.

Funding acquisition: Zhendong Wu, Rasmus Fensholt, Veiko Lehsten.

Methodology: Zhendong Wu, Niklas Boke-Olén.

Supervision: Rasmus Fensholt, Jonas Ardö, Lars Eklundh, Veiko Lehsten.

Writing – original draft: Zhendong Wu.

Writing – review & editing: Zhendong Wu, Niklas Boke-Olén, Rasmus Fensholt, Jonas Ardö, Veiko Lehsten.

References

1. Kowalczyk E, Stevens L, Law R, Dix M, Wang Y, Harman I, et al. The land surface model component of ACCESS: description and impact on the simulated surface climatology. Aust Meteorol Oceanogr J. 2013; 63(1):65–82.

2. Krinner G, Viovy N, de Noblet-Ducoudre N, Ogee J, Polcher J, Friedlingstein P, et al. A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. Glob Biogeochem Cycles. 2005; 19(1). https://doi.org/10.1029/2003GB002199

3. Smith B, Warlind D, Arneth A, Hickler T, Leadley P, Siltberg J, et al. Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model. Biogeosciences. 2014; 11:2027–54.

4. Zaehle S, Friend A. Carbon and nitrogen cycle dynamics in the O-CN land surface model: 1. Model description, site-scale evaluation, and sensitivity to parameter estimates. Glob Biogeochem Cycles. 2010; 24(1). https://doi.org/10.1029/2009GB003522

5. Jones C, Robertson E, Arora V, Friedlingstein P, Shevliakova E, Bopp L, et al. Twenty-first-century compatible CO2 emissions and airborne fraction simulated by CMIP5 earth system models under four representative concentration pathways. J Clim. 2013; 26(13):4398–413.

6. Moss RH, Edmonds JA, Hibbard KA, Manning MR, Rose SK, Van Vuuren DP, et al. The next generation of scenarios for climate change research and assessment. Nature. 2010; 463(7282):747–56. https://doi.org/10.1038/nature08823 PMID: 20148028

7. Prentice IC, Bondeau A, Cramer W, Harrison SP, Hickler T, Lucht W, et al. Dynamic global vegetation modeling: quantifying terrestrial ecosystem responses to large-scale environmental change. Terrestrial ecosystems in a changing world: Springer; 2007. p. 175–92.

8. Scheiter S, Langan L, Higgins SI. Next-generation dynamic global vegetation models: learning from community ecology. New Phytol. 2013; 198(3):957–69. https://doi.org/10.1111/nph.12210 PMID: 23496172

9. Sitch S, Huntingford C, Gedney N, Levy P, Lomas M, Piao S, et al. Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five Dynamic Global Vegetation Models (DGVMs). Glob Change Biol. 2008; 14(9):2015–39.

10. Barman R, Jain AK, Liang M. Climate-driven uncertainties in modeling terrestrial gross primary production: a site level to global-scale analysis. Glob Change Biol. 2014; 20(5):1394–411.

11. Jung M, Vetter M, Herold M, Churkina G, Reichstein M, Zaehle S, et al. Uncertainties of modeling gross primary productivity over Europe: A systematic study on the effects of using different drivers and terrestrial biosphere models. Glob Biogeochem Cycles. 2007; 21(4). https://doi.org/10.1029/2006gb002915

12. Wu Z, Ahlström A, Smith B, Ardö J, Eklundh L, Fensholt R, et al. Climate data induced uncertainty in model based estimations of terrestrial primary productivity. Environ Res Lett. 2017; 12(06):4013.

13. Poulter B, Frank D, Hodson E, Zimmermann N. Impacts of land cover and climate data selection on understanding terrestrial carbon dynamics and the CO 2 airborne fraction. Biogeosciences. 2011; 8 (8):2027–36.

14. Jones P, Harris I. University of East Anglia Climatic Research Unit, CRU TS3. 21: Climatic Research Unit (CRU) Time-Series (TS) Version 3.21 of High Resolution Gridded Data of Month-by-month Variation in Climate (Jan. 1901—Dec. 2012). NCAS British Atmospheric Data Centre. 2013. https://doi.org/10.5285/D0E1585D-3417-485F-87AE-4FCECF10A992

15. Peterson TC, Karl TR, Jamason PF, Knight R, Easterling DR. First difference method: Maximizing station density for the calculation of long-term global temperature change. J Geophys Res Atmos. 1998; 103(D20):25967–74.

16. Scherrer SC. Present-day interannual variability of surface climate in CMIP3 models and its relation to future warming. International Journal of Climatology. 2011; 31(10):1518–29.

17. Sheffield J, Goteti G, Wood EF. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. J Clim. 2006; 19(13):3088–111.

18. Weedon G, Gomes S, Viterbo P, Shuttleworth W, Blyth E, Österle H, et al. Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. J Hydrometeorol. 2011; 12(5):823–48.

19. Cox P, Jones C. Illuminating the modern dance of climate and CO2. Science. 2008; 321(5896):1642–4. https://doi.org/10.1126/science.1158907 PMID: 18801988

20. Sitch S, Smith B, Prentice IC, Arneth A, Bondeau A, Cramer W, et al. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. Glob Change Biol. 2003; 9(2):161–85.

21. Smith B, Prentice IC, Sykes MT. Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space. Glob Ecol Biogeogr. 2001; 10(6):621–37.

22. Battin TJ, Luyssaert S, Kaplan LA, Aufdenkampe AK, Richter A, Tranvik LJ. The boundless carbon cycle. Nat Geosci. 2009; 2(9):598–600.

23. Wramneby A, Smith B, Zaehle S, Sykes MT. Parameter uncertainties in the modelling of vegetation dynamics—effects on tree community structure and ecosystem functioning in European forest biomes. Ecol Model. 2008; 216(3):277–90.

24. Parton W, Scurlock J, Ojima D, Gilmanov T, Scholes R, Schimel DS, et al. Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide. Glob Biogeochem Cycles. 1993; 7(4):785–809.

25. Parton WJ, Hanson PJ, Swanston C, Torn M, Trumbore SE, Riley W, et al. ForCent model development and testing using the Enriched Background Isotope Study experiment. J Geophys Res. 2010; 115(G4). https://doi.org/10.1029/2009jg001193

26. Keeling R, Piper S, Bollenbacher A, Walker J. Atmospheric carbon dioxide record from Mauna Loa. ESS-DIVE (Environmental System Science Data Infrastructure for a Virtual Ecosystem); Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), 2009. https://doi.org/10.3334/CDIAC/atg.035

27. Lamarque J-F, Kyle GP, Meinshausen M, Riahi K, Smith SJ, van Vuuren DP, et al. Global and regional evolution of short-lived radiatively-active gases and aerosols in the Representative Concentration Pathways. Clim Change. 2011; 109(1–2):191–212.

28. Hurtt G, Chini LP, Frolking S, Betts R, Feddema J, Fischer G, et al. Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. Clim Change. 2011; 109(1–2):117–61.

29. Harris I, Jones P, Osborn T, Lister D. Updated high-resolution grids of monthly climatic observations–the CRU TS3. 10 Dataset. International Journal of Climatology. 2014; 34(3):623–42.

30. Wei Y, Liu S, Huntzinger D, Michalak A, Viovy N, Post W, et al. The North American carbon program multi-scale synthesis and terrestrial model intercomparison project–part 2: environmental driver data. Geosci Model Dev Discussions. 2013; 6(4):5375–422.

31. Dee D, Uppala S, Simmons A, Berrisford P, Poli P, Kobayashi S, et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Q J Royal Meteorol Soc. 2011; 137(656):553–97.

32. Kanamitsu M, Ebisuzaki W, Woollen J, Yang S-K, Hnilo J, Fiorino M, et al. Ncep-doe amip-ii reanalysis (r-2). Bull Am Meteorol Soc. 2002; 83(11):1631–43.

33. Zhang Y, Rossow WB, Lacis AA, Oinas V, Mishchenko MI. Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data. J Geophys Res Atmos. 2004; 109(D19).

34. Jung M, Reichstein M, Margolis HA, Cescatti A, Richardson AD, Arain MA, et al. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. J Geophys Res: Biogeosciences. 2011; 116(G3). https://doi.org/10.1029/2003jd004457

35. New M, Lister D, Hulme M, Makin I. A high-resolution data set of surface climate over global land areas. Clim Res. 2002; 21(1):1–25.

36. Schneider U, Fuchs T, Meyer-Christoffer A, Rudolf B. Global precipitation analysis products of the GPCC. Global Precipitation Climatology Centre (GPCC), DWD, Internet Publikation. 2008; 112.

37. Simmons A, Uppala S, Dee D, Kobayashi S. ERA-Interim: New ECMWF reanalysis products from 1989 onwards. ECMWF newsletter. 2007; 110(110):25–35.

38. Willmott CJ. On the validation of models. Phys geogr. 1981; 2(2):184–94.

39. Ahlström A, Raupach MR, Schurgers G, Smith B, Arneth A, Jung M, et al. The dominant role of semi-arid ecosystems in the trend and variability of the land CO2 sink. Science. 2015; 348(6237):895–9. https://doi.org/10.1126/science.aaa1668 PMID: 25999504

40. Schiffer R, Rossow WB. The International Satellite Cloud Climatology Project(ISCCP)- The first project of the World Climate Research Programme. Bull Am Meteorol Soc. 1983; 64:779–84.

41. Ahlström A, Schurgers G, Smith B. The large influence of climate model bias on terrestrial carbon cycle simulations. Environ Res Lett. 2017; 12(1):014004.

42. Chen M, Zhuang Q, He Y. An efficient method of estimating downward solar radiation based on the MODIS observations for the use of land surface modeling. Remote Sens. 2014; 6(8):7136–57.

43. McGuire A, Christensen T, Hayes D, Heroult A, Euskirchen E, Kimball J, et al. An assessment of the carbon balance of Arctic tundra: comparisons among observations, process models, and atmospheric inversions. Biogeosciences. 2012; 9(8):3185–204.

44. Murray-Tortarolo G, Anav A, Friedlingstein P, Sitch S, Piao S, Zhu Z, et al. Evaluation of land surface models in reproducing satellite-derived LAI over the high-latitude Northern Hemisphere. Part I: Uncoupled DGVMs. Remote Sens. 2013; 5(10):4819–38.

45. Sitch S, Friedlingstein P, Gruber N, Jones S, Murray-Tortarolo G, Ahlström A, et al. Recent trends and drivers of regional sources and sinks of carbon dioxide. Biogeosciences. 2015; 12(3):653–79.

46. Farquhar G, von Caemmerer Sv, Berry J. A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. Planta. 1980; 149(1):78–90. https://doi.org/10.1007/BF00386231 PMID: 24306196

47. Piao S, Sitch S, Ciais P, Friedlingstein P, Peylin P, Wang X, et al. Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO2 trends. Glob Change Biol. 2013; 19(7):2117–32.

48. Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, et al. The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc. 1996; 77(3):437–71.

49. Hungerford RD, Nemani RR, Running SW, Coughlan JC. MTCLIM: a mountain microclimate simulation model. 1989. PMID: A1989CT90500001.

50. Mlawer EJ, Taubman SJ, Brown PD, Iacono MJ, Clough SA. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. J Geophys Res Atmos. 1997; 102(D14):16663–82.

51. Chou M-D. A solar radiation model for use in climate studies. J Atmos Sci. 1992; 49(9):762–72.

52. Uppala SM, Kållberg P, Simmons A, Andrae U, Bechtold Vd, Fiorino M, et al. The ERA-40 re-analysis. Q J Royal Meteorol Soc. 2005; 131(612):2961–3012.

53. Weedon GP, Gomes S, Viterbo P, Österle H, Adam JC, Bellouin N, et al. The WATCH Forcing Data 1958–2001: A Meteorological Forcing Dataset For Land Surface-And Hydrological-Models. WATCH technical report; 2010. Rep. 22, 41 pp. [Available online at http://www.eu-watch.org/publications/technical-reports.]

54. NTSG. Numerical Terradynamic Simulation Group. The University of Montana, 32 Campus Drive, Missoula, MT 59812, USA. 2017. [Available online at http://www.ntsg.umt.edu/project/modis/mod17.php].

55. Anav A, Friedlingstein P, Beer C, Ciais P, Harper A, Jones C, et al. Spatiotemporal patterns of terrestrial gross primary production: A review. Rev Geophys. 2015; 53(3):785–818.

56. Ardö J. Comparison between remote sensing and a dynamic vegetation model for estimating terrestrial primary production of Africa. Carbon Balance Manag. 2015; 10(1):8.

57. Zhao M, Running SW, Nemani RR. Sensitivity of Moderate Resolution Imaging Spectroradiometer (MODIS) terrestrial primary production to the accuracy of meteorological reanalyses. J Geophys Res: Biogeosciences. 2006; 111(G1). https://doi.org/10.1029/2004jg000004

58. Medany M, Niang-Diop I, Nyong T, Tabo R, editors. Background paper on impacts, vulnerability and adaptation to climate change in Africa. UNFCCC Convention, Ghana; 2006. p. 21–23.

59. Trenberth KE, Dai A, Van Der Schrier G, Jones PD, Barichivich J, Briffa KR, et al. Global warming and changes in drought. Nat Clim Change. 2014; 4(1):17–22.

60. Huffman GJ, Adler RF, Bolvin DT, Gu GJ, Nelkin EJ, Bowman KP, et al. The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. J Hydrometeorol. 2007; 8(1):38–55. https://doi.org/10.1175/Jhm560.1

61. Papale D, Reichstein M, Aubinet M, Canfora E, Bernhofer C, Kutsch W, et al. Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. Biogeosciences. 2006; 3(4):571–83.

62. Moffat AM, Papale D, Reichstein M, Hollinger DY, Richardson AD, Barr AG, et al. Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes. Agric For Meteorol. 2007; 147 (3):209–32.

63. Lasslop G, Reichstein M, Papale D, Richardson AD, Arneth A, Barr A, et al. Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. Glob Change Biol. 2010; 16(1):187–208.

64. Reichstein M, Falge E, Baldocchi D, Papale D, Aubinet M, Berbigier P, et al. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. Glob Change Biol. 2005; 11(9):1424–39.

65. Legates DR, McCabe GJ. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour Res. 1999; 35(1):233–41.