MAPPING GLOBAL GROSS PRIMARY PRODUCTIVITY ON THE GOOGLE EARTH ENGINE PLATFORM – DEVELOPING AND APPLYING AN IMPROVED PROCESS-BASED ECOSYSTEM MODEL

By

HUA SHANG

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Ecology and Evolution

Written under the direction of

Ming Xu

And approved by

____

____

____

____

New Brunswick, New Jersey

JANUARY 2023

ABSTRACT OF THE DISSERTATION

Mapping global gross primary productivity on the Google Earth Engine platform – developing and applying an improved process-based ecosystem model

by HUA SHANG

Dissertation Director:

Ming Xu

The global carbon cycle has changed in response to climate change, and the effects of these changes, caused by anthropogenic factors such as the burning of fossil fuels and landscape alterations, are expected to be widespread. Terrestrial gross primary productivity (GPP), the largest component flux of the global carbon cycle, plays a significant role in connecting the global carbon and water cycles and the energy balance between the atmosphere, biosphere, hydrosphere and pedosphere. Despite the development of various approaches and models for estimating terrestrial GPP at different scales, large discrepancies and uncertainties remain in long-term global GPP simulations. Therefore, it is of great value and necessity to better understand and accurately estimate the spatial and temporal patterns of terrestrial GPP. In this dissertation, we improved the performance of global terrestrial GPP simulation by: 1) improving the solar radiation transfer model within a canopy by considering multiple scattering and radiation partitioning; 2) reconstructing satellite-based leaf area index (LAI) data to minimize biases and errors caused by cloud contaminations and composite technique; 3) using high performance computing of the Google Earth Engine (GEE) platform. We estimated global

terrestrial GPP at 0.25° spatial resolution and 3-hour temporal intervals using our integrated process-based ecosystem model from 2001 to 2020.

In Topic1, we evaluated the performance of five climate variables derived from a new reanalysis dataset - air temperature, precipitation, downward shortwave radiation, air pressure, and vapor pressure deficit (VPD) - against observations from 167 worldwide flux tower sites at both daily and annual scales. The results showed that all of the variables performed reliably, with the exception of precipitation, which had a tendency to be overestimated. In addition, we examined the temporal and spatial patterns of these variables from 2001 to2020. We found that global air temperature, solar radiation, VPD, and precipitation showed significantly increasing trends at rates of 0.7°C/decade, 3.1W/m^2/decade, 0.15KPa/decade, and 49.6mm/decade, respectively, while air pressure did not show any significant changes over this time period. The climate variables also showed different spatial variations at the global scale and their changes over the past decades were not homogenous in space. In addition to evaluating the climate variables, we also assessed the performance of reconstructing MODIS LAI products in 24 typical regions, which covered a range of major climate and vegetation types. The MODIS LAI datasets were affected by cloud contamination and composite techniques and did not perform well in areas with long-term continuous cloud cover, where LAI values were severely underestimated. We developed a new clean-up algorithm to improve the LAI data by including spatiotemporal correlations of neighboring pixels and applied double logistic functions to achieve continuous LAI time series. The results showed that most of the outliers were detected and removed, and the fitted double logistic curves well characterized the variations and patterns of annual LAI, reasonably captured the timing of vegetation phenology between growing and non-growing seasons, and retained the duration of peak within the growing season for both single vegetation cycle and double vegetation cycles.

In Topic 2, we found that the good performance of the empirical radiation partitioning approach indicated that it could be used to derive the two radiation components - direct and diffuse - when only total solar radiation information was available. Additional, the absorption fraction simulated by the two stream approach, which considered multiple scattering, was lower than that estimated by Beer's law regardless of the LAI and diffuse radiation fraction. The discrepancy in absorption fraction reached up to 73% in an overcast day. We further compared the performance of the Beer's law (BL) model, the two-stream big-leaf (TS-BL) model, and our integrated radiative transfer (RTM) model – the two-stream two-leaf (TS-TL) model - in simulating GPP and found that our TS-TL model reduced the RMSE and bias by up to 72% and 81% based on the BL model, and up to 63% and 75% based on the TS-BL model, respectively. Overall, our integrated RTM (TS-TL model) exhibited large improvements and robust performance in estimating GPP, especially in areas with a dense vegetation cover.

In Topic 3, we developed a comprehensive process-based ecosystem model, driven by new reanalysis climate data and satellite-based LAI data, to estimate global GPP by using different biochemical photosynthesis models for C3 and C4 plants on the GEE platform. The results were evaluated by comparing the simulated GPP to observations from 167 flux tower sites, and the modeled GPP estimates were highly correlated to the flux tower observations for all vegetation types at both half-hour and annual scales. The annual global terrestrial GPP simulated by our integrated model ranged from 118 PgC to 134 PgC, with an average of 128 PgC, during 2001-2020, and showed a significantly increasing trend with an average rate of 0.71 PgC/yr globally. When compared to recent GPP estimates and products, our simulated results were within a reasonable range of global terrestrial GPP estimations but had some discrepancies due to the different models, parameters, and driving data used to simulate GPP. In addition, the sensitivity analysis exhibited that our simulated GPP was most sensitive to the biophysiological parameters V_{cmax25} and LAI, highlighting the need for accurate biophysiological parameters at large scales.

Acknowledgements

First of all, I would like to thank my advisor Ming Xu for his support, guidance, and encouragement in my Ph.D. career. I was so lucky that I had the opportunity to come to Rutgers University and work with Ming in the Ecology and Evolution program. His advices were always inspirational, encouraging and supportive, and his continuous support and advice helped me grow to be a scientist. I shall never forget the research values and the dreams that he has passed to me.

I am grateful to my committee members, Peter Morin, Richard Lathrop, and Mary Whelan, for their guidance and feedback, which are irreplaceable for me to successfully finish this dissertation. I would like to offer special thanks to Marsha Morin and Shaneika Nelson, who have provided valuable assistants in navigating life at Rutgers. I would like to thank my colleagues, Fen Zhao, Shujian Wang, Baibing Ma, Ariel Kruger, and Sadiya Tijjani, for their friendship and research assistance.

I feel very fortunate to have been supported by my family and friends throughout my time at Rutgers. My parents, Huiqin Lu and Longjiang Shang, showed their selfless and infinity love and support to me, and I owe my successes to them. I am grateful that my husband Yue Gu was able to accompany with me, encourage me, and support me during my most difficult time. I am thankful to my daughter Emmy, whose innocent smiles and soft hugs give me great strength. Thanks to my cat Xiuzi, who sat on my keyboard and blocked my screen every time I was ready to give up.

What doesn't break me will eventually make me stronger. I would like to thank myself for hanging in there, overcoming all the difficulties, and accomplishing the research and dissertation.

Table of Contents

ABSTRACT OF THE DISSERTATION..................... ii
Acknowledgements vi
Table of Contents vii
List of Tables ix
List of Figures x
1. Introduction 1
1.1. Background 1
1.2. Estimating terrestrial gross primary productivity 4
1.2.1. Eddy covariance technique 4
1.2.2. Remote sensing-based models 5
1.2.3. Process-based models 6
1.3. Radiation transfer models 8
1.3.1. Based on light profiles 8
1.3.2. Based on canopy layers 9
1.4. Google earth engine platform 10
1.5. Research objectivities 12
2. Materials and methods 15
2.1. Ancillary Data 15
2.1.1. Reanalysis dataset 15
2.1.2. Global atmospheric CO_2 18
2.1.3. Remote sensing datasets 19
2.1.4. Global C3 and C4 distribution map 21
2.1.5. Flux tower data 22
2.2. Model description 23
2.2.1. Reconstructing LAI data 23
2.2.2. Integrated canopy radiative transfer model 26
2.2.3. Stomatal conductance 31
2.2.4. Photosynthesis 33
2.3. Statistical analysis 37
2.4. Sensitivity analysis 39
3. Investigating the variation of climate variables from GLDAS 2.1 and assessing the performance of fitting double logistic functions of LAI 40
3.1. Evaluating the climate variables from GLDAS 2.1 using flux tower observations 40

3.2. Characterizing the temporal and spatial variations of climate variables from GLDAS 2.1 46
3.3. Assessing the performance of reconstructing LAI........ 52
3.3.1. Single vegetation cycle 52
3.3.2. Double vegetation cycles 55
3.3.3. Performance of fitting in global scale 57
4. Integrating an improved two-stream canopy radiative transfer model 59
4.1. Evaluation of radiation partitioning 59
4.2. Comparisons of the two stream approach against Beer's law model on radiation absorption by canopy 61
4.3. Improvements of our integrated RTM on estimating GPP 62
5. Mapping global terrestrial gross primary productivity from local sites to global values using an improved process-based ecosystem model on Google Earth Engine Platform during 2001-202068
5.1. Comparison against flux tower sites 68
5.2. Temporal and spatial variations in global terrestrial GPP 72
5.3. Global estimates of terrestrial GPP 77
5.4. Sensitivity analysis on estimated GPP 78
6. Summary 81
6.1. Summary 81
6.2. Uncertainties and limitations 85
Appendices 86
References 99

List of Tables

Table 2.1 Climate variables of GLDAS 2.1 used in this study. 17

Table 2.2 Land cover type according to annual University of Maryland (UMD) classification. . 20

Table 2.3 Parameters for stomatal conductance and photosynthesis……..……………………….32

Table 2.4 Temperature dependence parameters………………………………………………….36

Table 2.5 Photosynthetic parameters for V_{cmax}^{25}………………………………………………..37

Table 3.1 The RMSE and bias of the five climate variables between GLDAS 2.1 and flux tower sites………………………………………………………………………………………… 45

Table 3.2 Detailed information of the selected sample sites with single vegetation cycle………...52

Table 3.3 Performance of fitting double logistic function to the LAI data after removing noise for the sample sites with single vegetation cycle………………………………………………..54

Table 3.4 Detailed information of the selected sample sites with double vegetation cycles……....55

Table 3.5 Performance of fitting double logistic function to the LAI data after removing noise for the sample sites with double vegetation cycles……………………………………………...57

Table 4.1 Detailed information of the flux tower sites used to evaluate the radiative transfer model……………………………………………………………………………………..…63

Table 4.2 The RMSE (gC/m^2/day), Bias (gC/m^2/day), and R^2 between the GPP derived using Beer's law (BL), two stream approach with big-leaf model (TS-BL), and two stream approach with two-leaf model (TS-TL) and the GPP measurements from flux tower sites………………....64

Table 5.1 Annual GPP (KgC/m^2/yr) among different vegetation types during 2001-2020. (ENF: evergreen needleleaf forests, EBF: evergreen broadleaf forests, DNF: deciduous needleleaf forests, DBF: deciduous broadleaf forests, MF: mixed forests, CSH: closed shrublands, OSH: open shrublands, WSA: woody savannas, SAV: savannas, GRA: grasslands, WET: wetlands, CRO: croplands)……………………………………………………………………………………...74

Table A.1 Information of flux tower stations used for evaluating the climate variables derived from GLDAS 2.1… ……………………………………………………………………….............86

Table A.2 Information of flux tower stations used for evaluating radiation partitioning…..……..91

Table A.3 Information of flux tower sites used for GPP validation…………………….………..92

List of Figures

Figure 1.1 Flowchart of the major steps in this research. 13

Figure 2.1 Trends in global monthly averaged atmospheric CO_2.. 18

Figure 2.2 Global distribution of C3 and C4 vegetation, presented by the percentages of C4 (%)……………………………………………………………………………………………..22

Figure 2.3 Illustration of the fitted double logistic curves for (a) single vegetation cycle and (b) double vegetation cycles, and the corresponding parameters……………………………….……25

Figure 2.4 Overview of the integrated radiation transfer model that quantify the radiation absorbed by the canopy, which includes partitioning of radiation (yellow), canopy radiative transfer model (grey), and two-leaf model (green)…………………………………………………………26

Figure 2.5 Illustration of the radiation transfer through a canopy for diffuse (a) and direct (b) radiation… …………………………………………………………………………………30

Figure 3.1 Comparison of daily air temperature (°C) between the observations from flux towers and the data retrieved from GLDAS 2.1. 42

Figure 3.2 Comparison of daily downward shortwave radiation (W/m^2) between the observations from flux towers and the data retrieved from GLDAS 2.1………………………………………42

Figure 3.3 Comparison of daily air pressure (KPa) between the observations from flux towers and the data retrieved from GLDAS 2.1…………………………………………………………43

Figure 3.4 Comparison of daily VPD (KPa) between the observations from flux towers and the data retrieved from GLDAS 2.1………………………………………………………………43

Figure 3.5 Comparison of daily precipitation (mm) between the observations from flux towers and the data retrieved from GLDAS 2.1…………………………………………………………44

Figure 3.6 Comparison of annual means of a) air temperature, b) downward shortwave radiation, c) air pressure, d) vapor pressure deficit (VPD), and e) precipitation amount between the observations from flux towers and the data retrieved from GLDAS 2.1…………………………45

Figure 3.7 Time series of global annual a) air temperature, b) downward shortwave radiation, c) air pressure, d) vapor pressure deficit (VPD), and e) precipitation amount retrieved from GLDAS 2.1 from 2001 to 2020. The dashed lines indicate the annual trends for the corresponding climate variables…………………………………………………………………………………48

Figure 3.8 Spatial distributions of global annual average of a1) air temperature, b1) downward shortwave radiation, c1) air pressure, d1) VPD, and e1) precipitation and global annual trends (a2

- e2) during the period of 2001-2020. For the global annual trend maps, the colors showed the regions where trends were statistical significant at 5% significant level, while the trends in the regions with white color were not significant ($p > 0.05$)……………………………………………51
Figure 3.9 The fitted curves (green lines) using double logistic function after removing noise (red dots) of the LAI data derived from MODIS (white dots) for the 16 sample sites with single vegetation cycle in 2003. The values of LAI should be scaled by 0.1……………………………53
Figure 3.10 The fitted curves (green lines) using double logistic function after removing noise (red dots) of the LAI data derived from MODIS (white dots) for the 8 sample sites with double vegetation cycles in 2003. The values of LAI should be scaled by 0.1…………………………..56
Figure 3.11 Performance of fitting double logistic function of LAI in global scale in 2003…….58
Figure 4.1 Comparison of the observed and modeled diffuse radiation (W/m^2) among the 36 flux tower sites. 60
Figure 4.2 Comparison of the absorbed fraction of solar radiation estimated by Beer's law (solid line) and two stream approach (shaped lines) for vegetation canopy with leaf area index ranging from 0.1 to 8 and diffuse radiation fraction (f_b) of 0 (diamond), 0.2 (star), 0.5 (triangle), 0.8 (square), and 1 (circle), respectively. The solar zenith angle (SZA) was set to 30°……..…………62
Figure 4.3 Comparisons of the modeled GPP derived by Beer's law (orange), two stream approach with big-leaf model (green), and two stream approach with two-leaf model (red) and the GPP obtained from flux tower observations (black) for sites US-Wcr (a1, a2), FR-PUE (b1, b2), CA-Obs (c1, c2), CA-NS6 (d1, d2), SN-Dhr (e1, e2), and CN-Dan (f1, f2). The vegetation types for US-Wcr, FR-PUE, CA-Obs, CA-NS6, SN-Dhr, and CN-Dan are DBF, EBF, ENF, SHB, SAV, and GRA, respectively. The black lines in a2-f2 are the 1:1 lines………………………………65
Figure 4.4 Boxplots of GPP differences between the simulated GPP derived by Beer's law (red), two stream approach with big-leaf model (blue), and two stream approach with two-leaf model (black) and the GPP measurements from flux tower sites………………………………………67
Figure 5.1 Comparison of our model-simulated GPP against 167 flux tower-based GPP…......... 70
Figure 5.2 Correlation coefficient, RMSE (gC/m^2/day), and bias (gC/m^2/day) between our model-simulated and flux tower-based GPP for different vegetation types……………………………..71
Figure 5.3 Comparison of our model-simulated annual mean GPP and flux tower-based annual GPP. The black line is the regression line, and the red line is the 1:1 line. Unit: gC/m^2/day……72

Figure 5.4 Time series of global annual GPP from our study and from a set of previous estimates. The right table shows the corresponding linear regressions……………………………………73
Figure 5.5 Spatial distributions of global a) annual mean GPP and b) significant trend ($p < 0.05$) of GPP during the period of 2001-2020. (Unit: gC/m^2/yr)…………………………………….76
Figure 5.6 Difference map of annual mean GPP between our study and MOD17 for the period of 2001-2020. (Unit: gC/m^2/yr)………………………………………………………………….…78
Figure 5.7 Sensitivity analysis of our model. Each of the six variables (LAI: leaf area index, CO_2: atmospheric CO_2 concentration, Rad: downward shortwave radiation, Temp: air temperature, V_{cmax}^{25}: maximum rate of carboxylation at 25 degree C, and VPD: vapor pressure deficit) was changed ±30% with keeping everything else the same, and the GPP (gC/m^2/yr) results were compared………………………………………………………………………………………79

1. Introduction

1.1. Background

Global carbon cycle has changed in response to climate change, and the effects of these changes caused by anthropogenic factors such as the burning of fossil fuels and landscape alterations are expected to be widespread (Kondratyev et al. 2003, Dixon and Turner 1991). Global climate change and the increasing atmospheric CO_2 concentration have highlighted the importance of better understanding the global carbon cycle (Nemani et al. 2003, Zhang et al. 2014). Plants utilize solar radiation, CO_2, and water through photosynthesis for their growth and maintenance under changing environmental conditions such as temperature and nutrition. Simulating vegetation photosynthesis activities at different temporal and spatial scales can help to address global carbon budget issues, accurately predict future climate changes, and scientifically understand the crucial role of terrestrial ecosystems in supporting the sustainable development of human society. Terrestrial gross primary productivity (GPP), the amount of carbon fixed by plants during photosynthesis, is the largest component flux of the global carbon cycle and plays a significant role in connecting the global carbon, energy, and water cycles throughout the atmosphere, biosphere, hydrosphere and pedosphere (Cramer et al., 2001, Wu et al. 2010, Yan et al. 2017). Despite the development of various approaches and models of estimating terrestrial GPP at different scales, large discrepancies and uncertainties remain in long-term global GPP products. Therefore, it is of great value and necessity to better understand and accurately simulate the spatial and temporal patterns of terrestrial GPP, which is crucial for optimizing the estimation of global carbon sources and sinks.

Solar radiation, as the primary energy source for ecosystems, drives the biological processes such as photosynthesis and evapotranspiration and the exchanges of energy and mass between the atmosphere, vegetation, and soil (Song and Band 2004, Ligot et al. 2014). Consequently, the amount of energy plants can intercept greatly affects terrestrial production and the carbon cycle, and accurate modeling of photosynthetically active radiation absorbed by vegetation is the key to terrestrial GPP estimation (Alton et al. 2007). Therefore, simulating radiative transfer processes through the canopy is essential in ecosystem process models to understand how radiation is distributed within the canopy and how much radiation is absorbed by plants (Nilson and Ross 1997, Yuan et al. 2014b, Yuan et al. 2017). A variety of canopy radiation transfer models have been developed to model the distribution and processes of incoming radiation within and below the canopy (Sellers 1985, Nijssen and Lettenmaier 1999, Dai et al. 2004). Simplified radiation transfer models often omit processes such as multiple scattering or radiation partitioning that have proven to cause large biases when estimating carbon fluxes, while the high computation burden and difficulty in acquiring data and parameters for more complicated models can hinder long-term GPP simulation at large scales. Therefore, an appropriate canopy radiation transfer model that integrates vital processes is important in simulating long-term global terrestrial GPP. Furthermore, the quality and continuity of model input data is a key factor in accurately simulating global terrestrial GPP. The development of remote sensing technology provides the possibility and opportunity to investigate global carbon cycles at high spatial resolution, and many well-developed products have already been widely applied by climate and ecosystem models to simulate global GPP. However, noise and gaps in remote sensing products, caused by atmospheric conditions or the sensors themselves, bring greater uncertainty to GPP estimations. Leaf area index (LAI), a critical vegetation biological variable, is widely used to estimate ecosystem processes such as photosynthesis and evapotranspiration (Xiao et al. 2009). LAI frequently serves as a key input parameter for modeling the exchange of carbon, water, and energy between the terrestrial and atmosphere (Liu et al. 2012). It is common for ecosystem models to run in a steady state, and most current models do not consider disturbance or simplify the process at large scales. Satellite-based LAI, which reflects the actual real-time vegetation conditions, can help to solve this problem. However, noise points and gaps exist in LAI products due to the influence of cloud or snow cover, leading to large biases and errors in LAI datasets. For example, the MODIS LAI datasets suffer from cloud contaminations and the composite technique and do not work well in areas with long-term continuous cloud cover, such as the spring of the East Asian monsoon region where a single "Meiyu" rain event can last up to two months. Cloud contamination can severely underestimate LAI values. The discontinuity and inconsistency of LAI data in space and time directly affect the accuracy of ecosystem carbon cycle simulations (Yuan et al. 2011). Therefore, it is important to improve the quality of LAI products and obtain accurate and consistent estimates of LAI at a global scale.

Like LAI, climate variables, as key driving factors of biogeochemical processes, are crucial to global carbon cycle modelling and have been widely used as model inputs in most ecosystem process models (Smith et al. 1993, Reichstein et al. 2013, Yang et al. 2017). While traditional weather stations provide accurate and valuable information for local and regional research, their irregular distribution limits their ability to depict spatial variations at a global scale. Reanalysis datasets, such as ERA-Interim, GLDAS, and NCEP, which combine land surface models, ground observations, and satellite data, provide global land surface states and fluxes in near-real time at high spatial and temporal resolution (Bosilovich et al. 2008, Mooney et al. 2011). Evaluating the performance of reanalysis data as model input data and investigating the patterns and variations of long-term trends in key climate variables, such as air temperature, precipitation, and solar radiation, is necessary for a better understanding and modelling of the global carbon cycle.

1.2. Estimating terrestrial gross primary productivity

Direct measurement of GPP does not exist, but various approaches for estimating GPP at multiple scales have been developed in the past decades (Piao et al. 2013, Ma et al. 2015, Sun et al. 2019), including flux tower estimates via the eddy covariance technique, remote sensing-based models, and process-based models. These approaches have their own unique strengths and limitations in meeting different demands for understanding the global carbon cycle at different scales.

1.2.1. Eddy covariance technique

The eddy covariance method is a micrometeorological method that is based on the turbulent transport theory to directly observe the exchanges of gases, energy, and momentum between the atmosphere and biosphere without disturbing the ecosystem (Baldocchi et al. 1988). Nowadays, the eddy covariance has been developed as a key technique for measuring the exchanges of net ecosystem CO_2, water vapor, and energy fluxes, and it provides powerful data support for plant ecophysiological studies and modeling of water and carbon cycles at regional and global scales. Over the past thirty years, the eddy covariance technique has been widely used in different ecosystems including forests, grasslands, and croplands in Asia (Zhang et al. 2007, Xiao et al. 2013), Europe (Morales et al. 2005, Papale et al. 2015), and America (Amiro et al. 2006, Amiro et al. 2010, Liu et al. 2022).

GPP can be obtained from measurements of net ecosystem exchange (NEE) between the atmosphere and terrestrial ecosystems, which uses the eddy covariance technique based on flux towers (Reichstein et al. 2005, Lasslop et al. 2010, Anav et al. 2015). Hundreds of worldwide flux tower networks that cover a large range of climate and biome types can provide continuous estimates of GPP, which play a pivotal role in understanding local carbon cycles and act as validation and calibration for global carbon models (Baldocchi et al. 2001, Friend et al. 2007). While GPP cannot be directly measured, flux towers provide probably the best estimates of GPP fluxes at the ecosystem level and have been used as ground-truth observations in numerous studies to calibrate and evaluate different models.

However, the estimates from flux towers only represent the fluxes at the scale of the tower footprint, which ranges between hundred meters and kilometers depending on the homogeneity of the vegetation (Xiao et al. 2010). And retrieving large-scale GPP estimates by scaling up data from flux towers has many uncertainties and depends on the availability of sufficient data, especially for long-term extrapolation (Beer et al. 2010, Anav et al. 2015).

1.2.2. Remote sensing-based models

Remote sensing (RS) datasets have been widely used in various models to estimate GPP, and the approaches are typically based on light use efficiency models (Monteith 1972, Running et al. 2000, Yuan et al. 2007) or empirical relationships with vegetation indices (Running et al. 2004, Sims et al. 2008, Li et al. 2013). They are efficient at exploring the spatial and temporal dynamics of plant growth at large scales and have relatively straightforward expressions (Song et al. 2013, Yuan et al. 2014, Sun et al. 2018, Sun et al. 2019).

The light use efficiency models were developed on the basis of the concept of radiation conversion efficiency, and they assumes that GPP is directly associated with the absorbed photosynthetically active radiation and is substantially dependent on the environmental conditions and the maximum light use efficiency (Monteith, 1972). The general form of the light use efficiency model can be expressed as:

$$GPP = PAR \times fPAR \times \varepsilon_{max} \times f(T, VPD, \dots) \quad \text{(Eq. 1.1)}$$

where PAR is the incident photosynthetically active radiation, $fPAR$ is the fraction of PAR that vegetation canopy is absorbed, and ε_{max} is the maximum light use efficiency, which is adjusted by multiple environmental scalars, such as air temperature $f(T)$ and vapor pressure deficit $f(VPD)$. And the vegetation index-based empirical models suggest that GPP could be directly estimated through empirical relationships with spectral-related indexes (Noumonvi et al. 2019).

$$GPP = a \times VI + b \quad \text{(Eq. 1.2)}$$

where VI is the vegetation index, a and b are regression constants.

The RS-based models are characterized by their large spatial coverage, temporal consistency, and straightforward computation, so they have the potential to investigate the spatial and temporal patterns of carbon fluxes at both regional and global scales (Pei et al. 2022). However, large variability still exists in explaining the inter-annual variations in GPP using the RS-based models due to the limitation in modelling the underlying mechanisms, especially at the global scale (Keenan et al. 2012, Liu et al. 2014, Yuan et al. 2014, Yan et al. 2017, Zheng et al. 2020).

1.2.3. Process-based models

The process-based ecosystem models, which are based on principles of ecology, biophysiology, and geochemistry, are also frequently used to understand and predict the storage, flux, and circulation of carbon, water, and other mineral nutrients in terrestrial ecosystems. Considerable efforts have been made to develop process-based models, such as Century (Parton et al. 1993), Biome-BGC (Running and Hunt 1993), LPJ (Sitch et al. 2003), and CASA (Potter et al. 1993), to estimate terrestrial GPP (Moorcroft 2006, Liu et al. 2014, Prentice et al., 2014, Anav et al. 2015).

The primary physiological processes that are generally used to simulate carbon assimilating include photosynthesis, transpiration, canopy radiative transmission, and stomatal conductance. Farquhar et al. (1980) described a biochemical photosynthesis model at a leaf level, assuming that the CO_2 uptake rate is limited by either RuBP carboxylase (Rubisco) or RuBP regeneration and the enzymatic components are all temperature-dependent. Currently, the Farquhar's photosynthesis model is widely accepted as theoretical basis, and the corresponding equations are generally included into all process-based models for GPP simulations. At the meantime, additional physiological processes, such as CO_2 diffusion, stomatal conductance, and canopy radiative transfer, are coupled into Farquhar's photosynthesis model when estimating GPP. Upscaling to plant or ecosystem levels, vegetation canopy is commonly treated as one big leaf, two leaves (sunlit and shaded), or multiple layers for different demand in the process-based models. And an intact canopy radiative transfer model that describes the absorption, reflection, and scattering of light provide elaborate physical processes to measure carbon assimilation. Most of the commonly used process-based models are operated on annually, monthly, or at most daily scales. The non-uniform changes of meteorological elements, especially solar radiation, within a day might affect the simulation results and bring large biases, since most ecological processes are nonlinear processes. The process-based models have the advantages in taking the effects of various environmental regulations into account at large scale and improving the understanding of ecological processes and global carbon cycle under global change. (Sitch et al. 2003, Morales et al. 2005, Sitch et al. 2008, Piao et al. 2009, Ryu et al. 2011, Liu et al. 2014), however tedious preparation of input data and parameters hinder many scientists from overcoming the computational burden in investigating the carbon cycle at large regions with shorter time scales.

1.3. Radiation transfer models

1.3.1. Based on light profiles

To examine the light profile within the canopy, the processes of radiation transfer through canopy have been modeled using different approaches. Monsi and Saeki (1953) first introduced Beer's law into ecosystem to quantify the light attenuation through the canopy. According to Beer's law, solar radiation decreases exponentially with the increasing depth through canopy without considering scattering (Monteith and Unsworth 2013), and the equation can be described as

$$I = I_0 e^{-K_b L} \qquad \text{(Eq. 1.1)}$$

where I_0 and I are the radiation intensities arriving at the top of the canopy and penetrating below the canopy layer, respectively. L is the cumulative leaf area index measured downwards from the top of the canopy, and K_b is the extinction coefficient of the canopy. In many ecosystem process models, Beer's law is commonly coupled within the canopy radiative transfer model to estimate the absorption and transmission of solar radiation for investigating the photosynthesis or evapotranspiration processes under different light conditions (Running and Hunt 1993). Although Beer's law performs well in predicting the average conditions of radiation below the canopy, previous studies showed that it is inadequate to model the interception within canopy (Larsen and Kershaw 1996, Nijssen and Lettenmaier 1999). Additionally, multiple scattering, which may increase the radiation below the canopy by up to 100% has not been accounted in Beer's law theory (Nijssen and Lettenmaier 1999).

To overcome the limitations and improve the canopy radiative transfer model, a two-stream approximation for radiation transfer through the vegetation canopy that considered multiple scattering was developed by Dickinson (1983) and Sellers (1985). In this two stream radiation transfer model, the changes in upward and downward radiation streams in a deep canopy are expressed by two differential equations with the considering of reflection, transmission, and absorption, and the general equations can be expressed by

$$-\frac{dI^{+}}{dL} = -K_b I^{+} + K_b \frac{\alpha}{2} I^{+} + K_b \frac{\alpha}{2} I^{-} \qquad \text{(Eq. 1.2)}$$

$$\frac{dI^{-}}{dL} = -K_b I^{-} + K_b \frac{\alpha}{2} I^{+} + K_b \frac{\alpha}{2} I^{-} \qquad \text{(Eq. 1.3)}$$

where I^{+} and I^{-} are the upward and downward radiation intensities within the canopy, and α is the leaf scattering coefficient. It assumes the radiation is scattered equally in the upward and downward directions in the canopy (Monteith and Unsworth 2013). Recently, Mahat and Tarboton (2012) extended this two-stream model from infinitely deep canopy to a finite canopy by using recursive superposition to obtain a solution, and the improved model could be applied to for both direct and diffuse radiation. Since the direct and diffuse radiation differ in the way they transfer through plant canopies and have different impacts on the nonlinear process of photosynthesis (de Pury and Farquhar 1997), it is now generally accepted that separately considering the transfer processes of direct and diffuse radiation improves the accuracy of modeling canopy radiation transfer processes. For example, previous studies found that an increased proportion of diffuse radiation leads to a higher light use efficiency and enhances vegetation photosynthesis (Gu et al. 2002, Alton et al. 2007).

1.3.2. Based on canopy layers

Three major groups of canopy radiation models include the big leaf model (Amthor 1994, Lloyd et al. 1995, Sellers et al. 1996), two-leaf model (Norman 1980, de Pury and Farquhar 1997, Wang and Leuning 1998, Dai et al. 2004), and multilayer model (Norman 1982, Leuning et al. 1995). The big leaf models, which treat the whole canopy as one big layer that retains all the properties of individual leaves, have been extensively used in many early studies (Amthor 1994, Lloyd et al. 1995, Sellers et al. 1996). They usually require fewer parameters and are relatively easier to test by field data. However, such models usually overestimate the photosynthesis due to the complex canopy structures (Amthor 1994, Dai et al.2004). To overcome the limitations of the big leaf model, multilayer models were developed that splitting the canopy into multiple layers and integrating the fluxes for each sub-layer to give the total flux for the whole canopy (Norman 1982, Leuning et al. 1995). The multilayer models consider the ecological processes of each layer inside the canopy in great details, such as leaf properties and leaf inclination angles, and they are regarded as the most accurate way to upscale fluxes from leaf to canopy (Luo et al. 2018). However, their expensive computational demand, especially for large scale, drives the need to develop alterative models. Two-leaf models have been proposed as simplifications to multilayer models (de Pury and Farquhar 1997, Wang and Leuning 1998, Dai et al. 2004). These models separate the canopy into two groups: sunlit leaves and shaded leaves, where the photosynthesis of sunlit leaves that receiving both direct and diffuse radiation tends to be light saturated, while the photosynthesis of shaded leaves that only absorb diffuse radiation depend on the intercepted radiation (de Pury and Farquhar 1997, Luo et al. 2018). The two-leaf models give very similar estimation of canopy photosynthesis compared to the simulation from multilayer models, but with far fewer computation time (Wang and Leuning 1998). Therefore, the two-leaf models have been extensively used in various ecosystem process models.

1.4. Google earth engine platform

Google Earth Engine (GEE) is an innovative cloud-based computing platform that archives a massive geospatial data catalog, and it provides a significant advance in processing petabyte-scale datasets at various scales (Gorelick et al. 2017, Mutanga and Kumar 2019). There are more than 200 public datasets, including satellite images, land cover data, and climate data, that are archived in GEE. Besides that, new datasets are updating daily for public use and researchers can upload their own data to GEE for different projects. Compared to other cloud computing platforms, GEE can avoid the tedious and time-consuming data downloading and uploading processes. In addition, GEE has powerful processing capacity and high-performance computing resources that automatically parallelizes the analysis on several CPUs across lots of computers in Google's data centers, which enables researchers to access and analyze mega-scale geospatial data (Gorelick et al. 2017). Moreover, the web-based application programming interface (API) of GEE is very user-friendly, and researchers can also choose to access the platform through Java script or Python API. There are many built-in functions that researchers can utilize for geospatial data processing and analyzing.

Although various well-developed models exist for estimating the terrestrial carbon cycle, data downloading and storage load, together with the huge computation cost, make it tremendously time consuming and even difficult to access for research at broad scales, especially with long time series (Ryu et al. 2011). Studies that expect high spatial and temporal resolutions need to deal with gigabytes or even terabytes of data, which might be the main problem that hinders the research progress. With all the strengths GEE possesses, researchers are able to obtain and analyze huge geospatial datasets for broader areas over long time periods, and have a rapid preview of the derived maps (Gorelick et al. 2017). In this study, we present our integrated process-based model built on the GEE platform to quantify the spatial and temporal patterns of the terrestrial carbon cycle at global scale without the technical and equipment burdens.

1.5. Research objectivities

The objectivities of this study are: 1) to build a comprehensive process-based ecosystem model that integrates the key physical and biogeochemical processes for simulating GPP on Google Earth Engine platform; 2) to increase the accuracy and performance of global terrestrial GPP simulation by evaluating the main input climate data that drives the ecosystem model, reconstructing the satellite-based LAI products, and improving the canopy radiative transfer model; 3) to achieve the terrestrial GPP estimations at global scale using our integrated process-based ecosystem model during the past two decades. And Figure 1.1 shows the flowchart of the major steps in my dissertation.

Three topics were introduced in this dissertation:

Topic 1: *Investigating the variation of climate variables from GLDAS 2.1 and assessing the performance of fitting double logistic functions of LAI*. In this topic, I aimed to reduce the biases from input data and increase the accuracy of Global GPP estimation by evaluating and improving the key model input data (Chapter 3).

Topic 2: *Integrating an improved two-stream canopy radiative transfer model*. In this topic, I aimed to improve the performance of simulating canopy radiation absorption and GPP by improving the radiative transfer model (Chapter 4).

Topic 3: *Mapping global terrestrial gross primary productivity from local sites to global values using an improved process-based ecosystem model on Google Earth Engine Platform during 2001-2020*. In this topic, I aimed to achieve the terrestrial GPP estimation at global scale by developing an improved process-based ecosystem model driven by satellite-based LAI data and reanalysis climate data during the past two decades using high performance computing (Chapter 5).

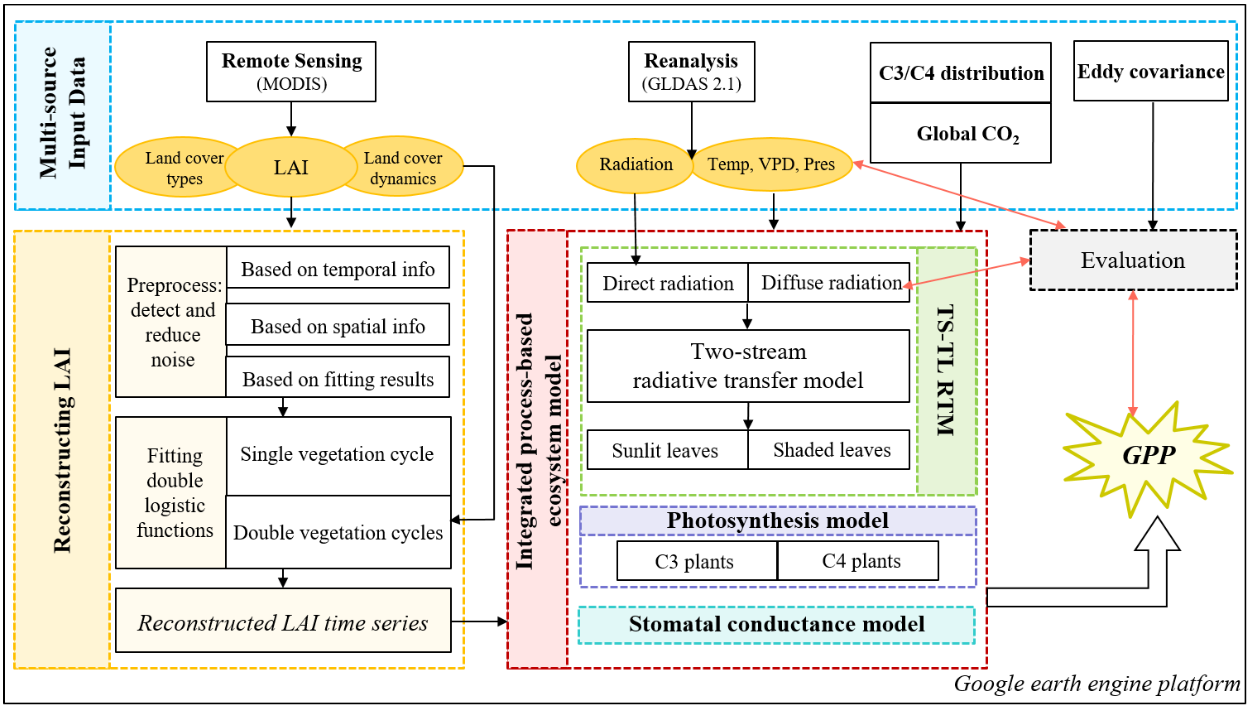


Figure 1.1 Flowchart of the major steps in this research.

My dissertation approaches the goals by the following steps:

(1) Topic 1 evaluated five GLDAS 2.1-derived climate variables that are essential for simulating global carbon cycle and are commonly used as input data to drive ecosystem models, including air temperature, precipitation, downward shortwave radiation, air pressure, and VPD (derived from specific humidity), against the observations from 120 worldwide flux tower sites. The temporal and spatial variations of the five climate variables from 2001 to 2020 were also investigated at global scale. Moreover, remote sensing LAI product (MCD15A3H) was reconstructed to obtain a high-quality and continuous time series by fitting double logistic functions after eliminating noise and outliers. Different double logistic functions were applied to grids with single vegetation cycle and grids with double vegetation cycles, and the corresponding fitting performance was discussed in this chapter.

(2) Topic 2 focused on improving the accuracy of modeling the radiation absorbed by the vegetation canopy and further increasing the performance of terrestrial GPP simulation by integrating a recently developed two stream radiative transfer model that considers multiple scattering in a finite canopy to a two-leaf model. In addition, an empirical radiation partitioning approach was evaluated against 258 site-years from 36 flux tower sites.

(3) In topic 3, a comprehensive process-based model that coupled the improved two stream radiation transfer model was developed on the Google Earth Engine platform, and the simulated GPP was evaluated against 167 flux tower sites. The spatial and temporal patterns and trends in global terrestrial GPP during 2001-2020 were examined for different vegetation types, and the comparisons of global GPP estimates from recent studies and products were discussed. In addition, the sensitivities of our model to environmental and biological drivers were also investigated in this chapter.

2. Materials and methods

2.1. Ancillary Data

2.1.1. Reanalysis dataset

In the past decades, data assimilation techniques that assimilate weather forecast information, ground observation data and remote sensing data into analysis products provide many global climate datasets with high spatial resolution for a long time period (Bosilovich et al. 2008, Mooney et al. 2011). Many well-known reanalysis datasets are commonly used for global climate and ecosystem modelling, such as the National Centers for Environmental Prediction and the National Center for Atmospheric Research reanalysis (NCEP/NCAR, Kalnay et al. 1996), the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Reanalysis (ERA-Interim, Dee et al. 2011), and the Japanese 55-year Reanalysis (JRA-55, Kobayashi et al. 2015). The Global Land Data Assimilation System (GLDAS) is a new generation of reanalysis dataset that is jointly developed by the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC) and the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction (NCEP). GLDAS integrated ground-based observations, remote sensing images, radar precipitation measurements, and outputs from numerical prediction models into advanced Land Surface Models (LSM) using data assimilation techniques to produce a global, high-resolution, offline (uncoupled to the atmosphere) terrestrial modeling system that simulates global land surface states and fluxes in near-real time (Rodell et al. 2004). GLDAS currently drives five models, including Mosaic, Noah, the Community Land Model (CLM), the Variable Infiltration Capacity model (VIC), and the Catchment Land Surface Model (CLSM), to produce a massive archive of global modeled and observed outputs from

1948 to present with spatial resolution of 1 degree and 0.25 degree and temporal resolution of 3-hourly, daily, and monthly. There are three components of GLDAS version 2: (1) GLDAS 2.0 is forced entirely with the Princeton Meteorological Forcing Dataset and provides a temporally consistent series (with 3- hourly, daily, and monthly temporal intervals) from 1948 through 2014; (2) GLDAS 2.1 is forced with a combination of model and observation data, and contains 3- hourly and monthly data spanning from 2000 to the present; (3) GLDAS 2.2 uses data assimilation (while the other two products are "open-loop") to produce daily data output from 2003 to the present.

In consideration of the temporal and spatial resolution, time coverage, and the data availability on the Google Earth Engine, GLDAS 2.1 (simulated by the Noah-3.6) was used in this study. The dataset contains 36 parameters with 3-hourly temporal interval, 0.25-degree spatial resolution, and spatial extent from -60° to 90° (latitude) and -180° to 180° (longitude) in the geographic coordinate system. We retrieved five variables from the dataset, including air temperature (K), total precipitation rate (kg/m^2/s), downward shortwave radiation (W/m^2), specific humidity (kg/kg), and air pressure (Pa), and the detailed information of the variables is listed in Table 2.1. The variable names with extension "_inst" are instantaneous variables, while those with extension "_tavg" are backward 3-hour averaged variables. The GLDAS 2.1 dataset is available to use on Google Earth Engine Platform ("NASA/GLDAS/V021/NOAH/G025/T3H", https://developers.google.com/earth-engine/datasets/catalog/NASA_GLDAS_V021_NOAH_G025_T3H).

Table 2.1 Climate variables of GLDAS 2.1 used in this study.

Variable name	Unit	Description	min	max	Spatial resolution	Temporal resolution
Tair_f_inst	K	Air temperature	206.8	327.66	0.25°	3-hourly
Rainf_f_tavg	$kg/m^2/s$	Total precipitation rate	0	0.01	0.25°	3-hourly
SWdown_f_tavg	W/m^2	Downward short-wave radiation flux	-56.93	30462.8	0.25°	3-hourly
Qair_f_inst	kg/kg	Specific humidity	-0.02	0.07	0.25°	3-hourly
Psurf_f_inst	Pa	Surface pressure	44063.1	108344	0.25°	3-hourly

We further calculated the vapor pressure deficit (VPD) as one of our input variables to the model by using specific humidity. The VPD is the difference between the amount of moisture in the air and how much moisture the air can hold when it is saturated (Howell and Dusek 1995). VPD as an important driver of atmospheric water demand for plants, influences terrestrial ecosystem function and photosynthesis (Rawson et al. 1977). VPD is commonly used in stomatal conductance models to predict leaf stomatal conductance and photosynthesis (Leuning 1995). Since most reanalysis climate datasets only provide dew point temperature or specific humidity instead of vapor pressure deficit (VPD), calculations were needed to get the VPD. The VPD (KPa) can be derived from the difference between saturated vapor pressure (es, KPa) and actual vapor pressure (ea, KPa) (Yoder et al. 2005):

$$es = 0.61078 \cdot e^{\frac{17.27 \cdot T}{T+237.3}} \quad \text{(Eq. 2.1)}$$

$$ea = 1.6077 \cdot \mathrm{q} \cdot \mathrm{P}_a \quad \text{(Eq. 2.2)}$$

$$VPD = es - ea \quad \text{(Eq. 2.3)}$$

where T is air temperature (°C), q is the specific humidity (kg/kg), and P_a is atmospheric pressure (KPa).

2.1.2. Global atmospheric CO_2

The Global Monitoring Laboratory (GML) of the NOAA/ESRL monitoring program provides high-precision measurements of the global atmospheric distribution and trends of greenhouse gases. The global averaged surface carbon dioxide from 1980 to the present are calculated based on 43 marine boundary layer (MBL) sampling sites from the NOAA/GML global air sampling network. The air samples collected from these sites are predominantly of well-mixed clean air to eliminate the influences from nearby sinks and sources of CO_2, such as vegetation and human activities (Masarie and Tans 1995). Here, the monthly global averaged CO_2 from 2001 to 2020 were downloaded and used (https://gml.noaa.gov/ccgg/trends/gl_data.html). Figure 2.1 shows the trends in global monthly averaged atmospheric CO_2.

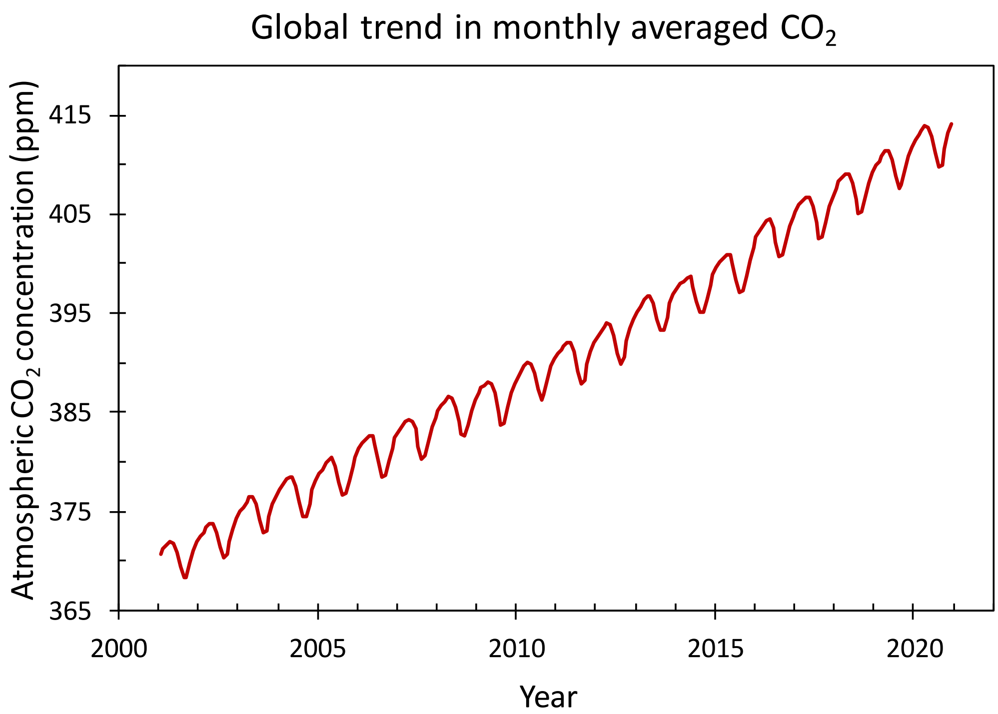


Figure 2.1 Trends in global monthly averaged atmospheric CO_2.

2.1.3. Remote sensing datasets

Remote sensing data is commonly used to improve our understanding of global dynamics and processes occurring on the land, in the oceans, and in the atmosphere. In this dissertation, remote sensing data was used to support and drive our model and evaluate the simulation results, and the four remote sensing datasets we used, including leaf area index (LAI), land cover type, land cover dynamics (global vegetation phenology), and gross primary productivity (GPP), are all from Moderate Resolution Imaging Spectroradiometer (MODIS) satellites.

LAI data is derived from the MCD15A3H Version 6.1 Level 4 product, which is a 4-day composite dataset with 500-meter spatial resolution spanning from 2002 to the present. LAI is defined as one-sided green leaf area per unit ground area in broadleaf canopies and one-half the total needle surface area per unit ground area in coniferous canopies in this product. The LAI dataset is available to use on the Google Earth Engine Platform ("MODIS/061/MCD15A3H", https://developers.google.com/earth-engine/datasets/catalog/MODIS_061_MCD15A3H).

Land cover type data is from the MCD12Q1 V6 product, and it provides global annual land cover types during 2001 to 2019 at 500-meter spatial scale. We used the Annual University of Maryland (UMD) classification system, which includes 16 different land cover types: evergreen needleleaf forests, evergreen broadleaf forests, deciduous needleleaf forests, deciduous broadleaf forests, mixed forests, closed shrublands, open shrublands, woody savannas, savannas, grasslands, croplands, permanent wetlands, urban and built-up lands, cropland/natural vegetation mosaics, non-vegetated lands, and water bodies (Table 2.2) , and the first 12 vegetation land types were used to perform the model. The land cover type dataset is available to use on the Google Earth Engine Platform ("MODIS/006/MCD12Q1", https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MCD12Q1).

Table 2.2 Land cover type according to annual University of Maryland (UMD) classification.

Land cover type	Description
Water Bodies	At least 60% of area is covered by permanent water bodies.
Evergreen Needleleaf Forests (ENF)	Dominated by evergreen conifer trees (canopy >2m). Tree cover >60%.
Evergreen Broadleaf Forests (EBF)	Dominated by evergreen broadleaf and palmate trees (canopy >2m). Tree cover >60%.
Deciduous Needleleaf Forests (DNF)	Dominated by deciduous needleleaf (larch) trees (canopy >2m). Tree cover >60%.
Deciduous Broadleaf Forests (DBF)	Dominated by deciduous broadleaf trees (canopy >2m). Tree cover >60%.
Mixed Forests (MF)	Dominated by neither deciduous nor evergreen (40-60% of each) tree type (canopy >2m). Tree cover >60%.
Closed Shrublands (CSH)	Dominated by woody perennials (1-2m height) >60% cover.
Open Shrublands (OSH)	Dominated by woody perennials (1-2m height) 10-60% cover.
Woody Savannas (WSA)	Tree cover 30-60% (canopy >2m).
Savannas (SAV)	Tree cover 10-30% (canopy >2m).
Grasslands (GRA)	Dominated by herbaceous annuals (<2m).
Permanent Wetlands (WET)	Permanently inundated lands with 30-60% water cover and >10% vegetated cover.
Croplands (CRO)	At least 60% of area is cultivated cropland.
Urban and Built-up Lands	At least 30% impervious surface area including building materials, asphalt and vehicles.
Cropland/Natural Vegetation Mosaics	Mosaics of small-scale cultivation 40-60% with natural tree, shrub, or herbaceous vegetation.
Non-Vegetated Lands	At least 60% of area is non-vegetated barren (sand, rock, soil) or permanent snow and ice with less than 10% vegetation.

The land cover dynamics product (global vegetation phenology) from the MCD12Q2 V6 product provides the estimates of the timing of vegetation phenology at global scales. The dataset includes the information of the onset of greenness, greenup midpoint, maturity, peak greenness, senescence, greendown midpoint, and dormancy over a vegetation cycle, and normally there are one or two observed vegetation cycles in a year. This product is yearly data from 2001 to 2018 with 500-meter spatial resolution. The land cover dynamics dataset is available to use on the Google Earth Engine Platform (“MODIS/006/MCD12Q2”, https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MCD12Q2).

The GPP dataset from the MOD17A2H V6 Gross Primary Productivity (GPP) product is a cumulative 8-day composite data with a 500-meter spatial resolution from 2000 to the present, and the GPP is calculated based on the light use efficiency model. The GPP dataset is available to use on Google Earth Engine Platform (“MODIS/006/MOD17A2H”, https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MOD17A2H).

2.1.4. Global C3 and C4 distribution map

The global distribution of C3 and C4 plants is crucial for accurately simulating the exchanges of carbon, water, and energy between atmosphere and biosphere due to the physiological and functional distinctions between C3 and C4 plants, such as physiological structures, photosynthetic pathways, and the responses to changing CO_2, light, and temperature. Therefore, we used the distribution map of global C3 and C4 vegetation at 1-degree spatial scale provided by Still et al. (2003) to incorporate different physiological processes of C3 and C4 vegetation. It was developed by combining remote sensing products, physiological modeling, global crop fractions, and national harvest area data, and as a result the C4 vegetation covers approximately 18.8 million km^2, the C3 vegetation covers about 87.4 million km^2, and the bare ground and ice cover the rest of the land surface (Still et al. 2003). Figure 2.2 shows the global percentages of C4 vegetation. We resampled the distribution map using nearest neighborhood method to a 0.25-degree resolution.

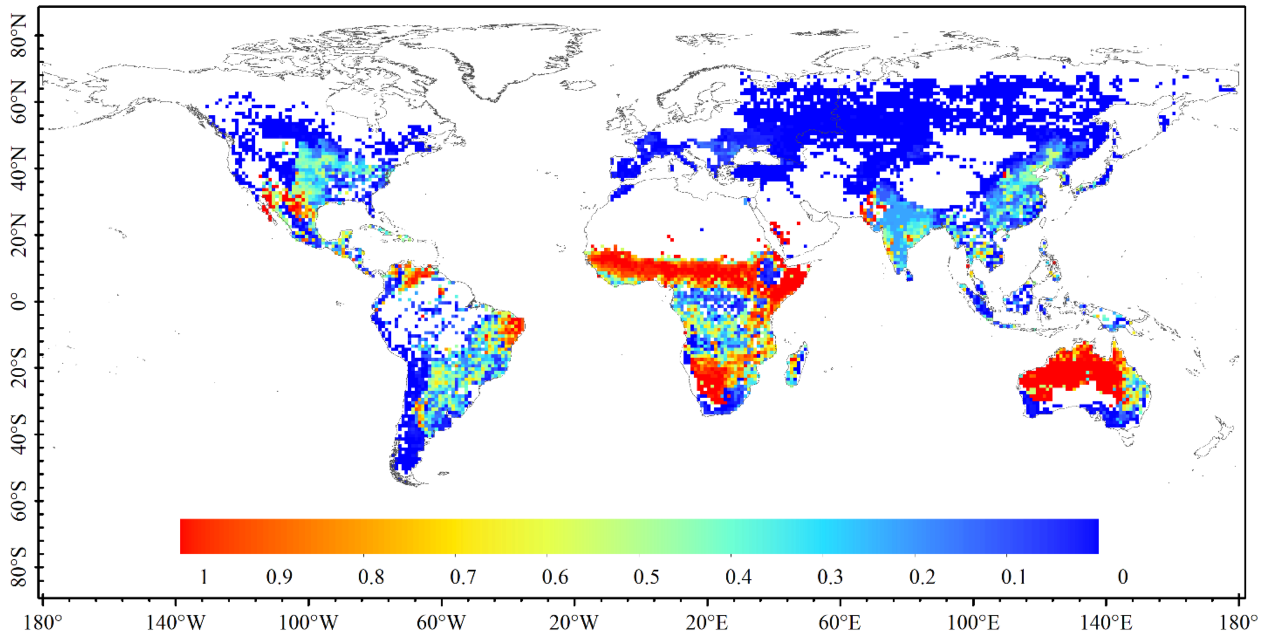


Figure 2.2 Global distribution of C3 and C4 vegetation, presented by the percentages of C4 (%).

2.1.5. Flux tower data

The FLUXNET is a global network that integrates multiple regional flux networks, such as AmeriFlux, AsiaFlux, ChinaFlux, EuroFlux, OzFlux, and FLUXNET-Canada (Baldocchi et al. 2001), and it is the most comprehensive platform for integration and sharing flux measurements currently. Globally, the FLUXNET synthesizes datasets from hundreds of observation sites that measure carbon, water, and energy exchanges between the atmosphere and biosphere based on eddy covariance methods.

The FLUXNET2015 Dataset, which is hosted by the Lawrence Berkeley National Laboratory, is the most recent FLUXNET data product after the FLUXNET Marconi Dataset (2000) and the FLUXNET LaThuile Dataset (2007). The dataset contains not only the carbon and energy fluxes but also the meteorological and biological measurements collected from 212 sites around the globe spanning from the early 1990s to 2014 (Pastorello et al. 2020). The tower sites cover 15 land cover types, including evergreen needleleaf forest (ENF), evergreen broadleaf forest (EBF), deciduous needleleaf forest (DNF), deciduous broadleaf forest (DBF), mixed forest (MF), grassland (GRA), cropland (CRO), closed shrubland (CSH), open shrubland (OSH), wetland (WET), savanna (SAV), woody savanna (WSA), snow or ice covered land (SNO), urban and built-up land, and barren or sparsely vegetated land. The outputs of FLUXNET2015 include over 200 variables, such as gross primary productivity, ecosystem respiration, net ecosystem exchange, soil heat flux, sensible heat, latent heat, air temperature, soil temperature, longwave radiation, and shortwave radiation, and provides five major temporal resolutions including half-hourly/hourly, daily, weekly, monthly, and yearly. The dataset is now available to download for public (https://fluxnet.org/data/fluxnet2015-dataset/).

2.2. Model description

2.2.1. Reconstructing LAI data

The 4-day LAI data derived from MODIS MCD15A3H product was reconstructed to obtain continuous time series by fitting the double logistic function, and the noise and outliers were detected and removed before the fitting. The first processing step was to eliminate the noise and outliers based on the spatiotemporal correlations of neighboring pixels. We first checked the multi-year time series of each pixel, and removed the outliers based on residual analysis and boxplot. The outliers were defined either larger than the maximum or smaller than the minimum (Eq. 2.4-2.5). Then we applied a 5km by 5km moving window spatially to detect and remove the outliers based on boxplot.

$$Minimum = P_{25} - (P_{75} - P_{25}) \times 1.5 \qquad \text{(Eq. 2.4)}$$

$$Maximum = P_{75} + (P_{75} - P_{25}) \times 1.5 \qquad \text{(Eq. 2.5)}$$

where P_{25} is the 25th percentile and P_{75} is the 75th percentile. After that, double logistic function was fitted to the dataset after removing the apparent outliers from the LAI series. We then removed the noise in a third time using boxplot based on the differences between original and predicted LAI values.

After all the preprocessing steps, double logistic functions were fitted to the cleaned data to obtain the continuous curves of LAI series, which is applicable to drive the global carbon cycle model as major input. Double logistic function have been applied on the EVI, NDVI, or LAI time series to obtain phenology information in previous studies (Cai et al. 2017, Testa et al. 2018). The two sigmoid curves could well indicate green-up and senescence phases of vegetation growth. When fitting the double logistic function, we classified the pixels into two categories: grids with single vegetation cycle and grids with double vegetation cycles according to the numbers of vegetation cycles derived from the MODIS land cover dynamics product, and applied corresponding double logistic functions to the two categories. For the grids with a single vegetation cycle, the double logistic function we used in this study was described as (Gonsamo et al. 2012):

$$LAI(t) = \alpha_1 + \frac{\alpha_2}{1+e^{-\gamma_1(t-\beta_1)}} - \frac{\alpha_2}{1+e^{-\gamma_2(t-\beta_2)}} \qquad \text{(Eq. 2.6)}$$

where t is the day of year (DOY), $LAI(t)$ is the observed LAI at time t, α_1 is the background LAI, $\alpha_2 - \alpha_1$ is the difference between the background and the growth season plateau, γ_1 and γ_2 are the transition in slope coefficients, and β_1 and β_2 are the midpoints in DOY of these transitions for green-up and senescence/abscission, respectively.

For the grids with double vegetation cycles, we used the phenology information from MODIS land cover dynamics to detect the boundary between the two vegetation cycles. Then we fitted the double logistic function (Eq. 2.7) for each cycle and found the solution of the two functions.

$$LAI(t) = \alpha_1 + \frac{\alpha_2}{1+e^{-\gamma_1(t-\beta_1)}} - \frac{\alpha_3}{1+e^{-\gamma_2(t-\beta_2)}} \quad \text{(Eq. 2.7)}$$

where $\alpha_2 - \alpha_1$ is the difference between the background and the amplitude of spring and early summer plateau, and $\alpha_3 - \alpha_1$ is the difference between the background and the amplitude of late summer plateau and autumn. Figure 2.3 shows the illustration of the fitted curves and corresponding parameters. After the fitting, the 7 derived parameters could provide the LAI values for any given time.

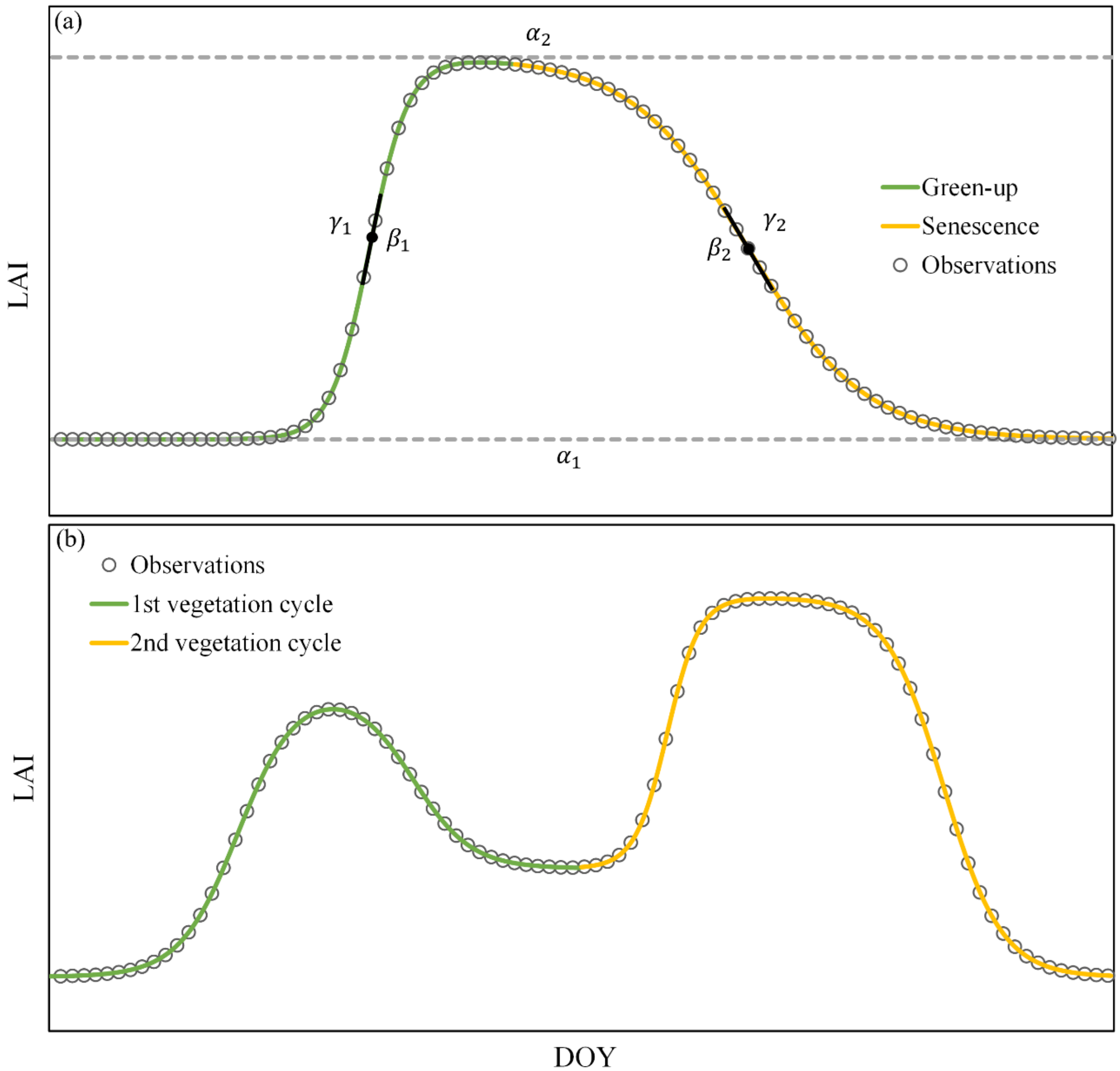


Figure 2.3 Illustration of the fitted double logistic curves for (a) single vegetation cycle and (b) double vegetation cycles, and the corresponding parameters.

2.2.2. Integrated canopy radiative transfer model

First, we partitioned the total incoming radiation into direct and diffuse components based on the cloudiness fraction. Then we improved the two-stream radiative transfer approximation that considering multiple scattering within a finite canopy by using different scattering coefficient for direct and diffuse radiation and quantified the transmission and reflection factors for direct and diffuse radiation, respectively. And finally, we coupled the improved model to a two-leaf model that considers the differences in the absorption of radiation between sunlit and shaded leaves. Figure 2.4 illustrates the overview of our two-leaf canopy radiative transfer models.

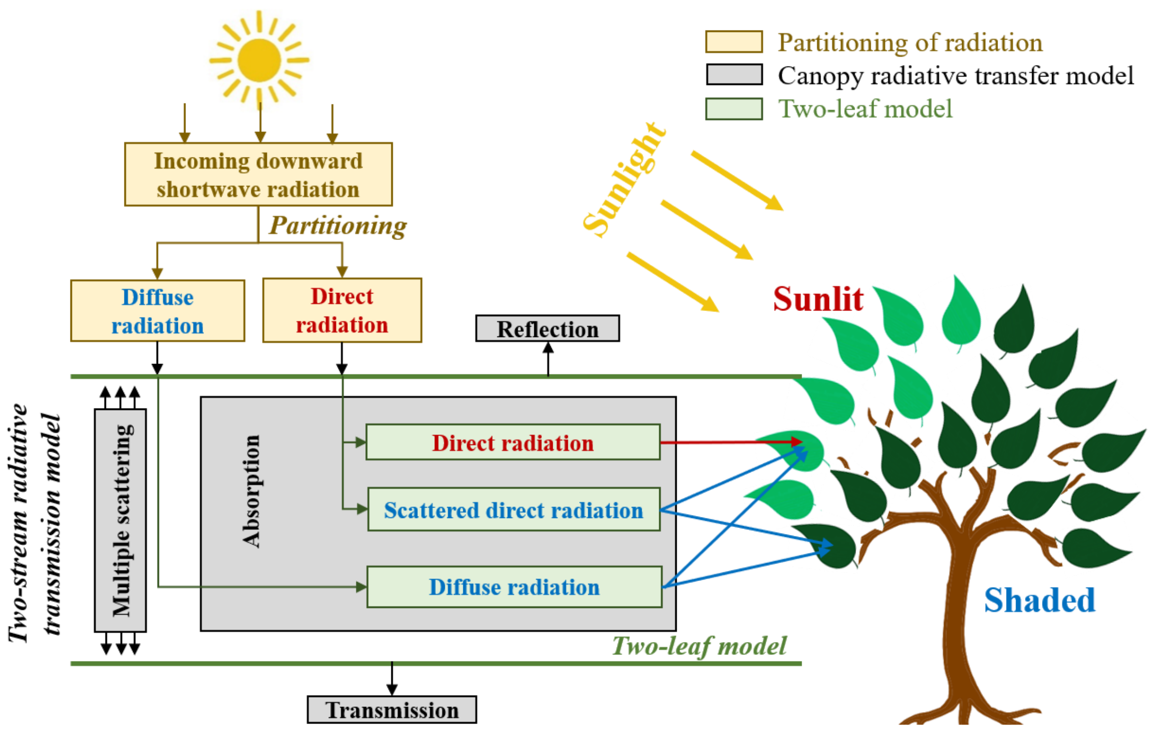


Figure 2.4 Overview of the integrated radiation transfer model that quantify the radiation absorbed by the canopy, which includes partitioning of radiation (yellow), canopy radiative transfer model (grey), and two-leaf model (green).

Partitioning of downward shortwave radiation to direct and diffuse radiation

The incoming direct radiation (S_b) and diffuse radiation (S_d) are partitioned from the total downward shortwave radiation (S_t) that reaching to the canopy according to Mahat and Tarboton

(2012). First, we separated the total fraction of radiation reaching to the canopy in reaching to the top atmosphere (AT, S_t/S_0) to direct (AT_b) and diffuse (AT_d) components. These two components can be calculated as:

$$AT_b = \lambda\, AT_c\left(1 - C_f\right) \qquad \text{(Eq. 2.8)}$$

$$AT_d = AT - AT_b \qquad \text{(Eq. 2.9)}$$

where λ is the ratio of direct to total radiation for clear sky, which assumes a fraction λ of AT is direct under a clear sky and all the radiation is diffuse when the sky is fully overcast, and the value of 6/7 was used in our model, AT_c represents the clear sky transmission factor and equals max($AT, a_s + b_s$), a_s is the fraction of extraterrestrial radiation on overcast days, $a_s + b_s$ is the fraction of extraterrestrial radiation on clear days, and a_s = 0.25 and b_s = 0.50 are recommended by Shuttleworth (1993). The cloudiness fraction C_f is assumed to be 0 on a clear sky while equals to 1 on a fully cloudy sky where all the radiation is from diffuse radiation. We estimated the cloudiness fraction based on the total incoming shortwave radiation using the following empirical relationship:

$$S_t = \left(a_s + b_s\left(1 - C_f\right)\right) S_0 \qquad \text{(Eq. 2.10)}$$

where S_0 is the extraterrestrial radiation calculated by $S_c \cos\theta$ (S_c as solar constant, approximately equals to 1367 W/m^2, and θ as solar zenith angle).

After we have the AT_b and AT_d following the above equations, then the direct radiation and diffuse radiation are given by:

$$S_b = \frac{AT_b}{AT} S_t \qquad \text{(Eq. 2.11)}$$

$$S_d = \frac{AT_d}{AT} S_t \qquad \text{(Eq. 2.12)}$$

Canopy radiative transfer model

Once the direct radiation and diffuse radiation are separated from the total downward shortwave radiation, we assumed that the two components penetrate the canopy separately. A radiative transmission model considering multiple scattering using a two-stream approach in a finite canopy (Mahat and Tarboton 2012) was applied in this paper. This model was developed based on Beer's law but was adjusted for multiple scattering and reflection. And it assumed the radiation is scattered equally in an upward and downward direction and the scattering direction is along the same path as the incoming radiation. This model considered that the incoming radiation is either transmitted through the canopy, or reflected by the canopy, or absorbed by the canopy.

The transmission factor (τ) and reflection factor (β) with multiple scattering for both direct and diffuse radiation could be estimated by Eq. 2.13 and Eq. 2.14:

$$\tau = \frac{\tau'[1-(\beta')^2]}{1-(\beta')^2(\tau')^2} \qquad \text{(Eq. 2.13)}$$

$$\beta = \frac{\beta'[1-(\tau')^2]}{1-(\beta')^2(\tau')^2} \qquad \text{(Eq. 2.14)}$$

The above equations for a finite canopy were obtained by recursive superposition of the solution for infinitely deep canopy. And τ' and β' are the corresponding transmission and reflection factors for an infinitely deep canopy, which could be calculated by:

$$\tau_b{}' = e^{-\sqrt{1-\alpha}K_b LAI} \qquad \text{(Eq. 2.15)}$$

$$\tau_d{}' = [(1-\sqrt{1-\alpha}GLAI)e^{-\sqrt{1-\alpha}GLAI} + \left(\sqrt{1-\alpha}GLAI\right)^2 E_i(1, \sqrt{1-\alpha}GLAI) \qquad \text{(Eq. 2.16)}$$

$$\beta' = \frac{1-\sqrt{1-\alpha}}{1+\sqrt{1-\alpha}} \qquad \text{(Eq. 2.17)}$$

where α is the leaf scattering coefficient (different values are used for direct and diffuse radiation, where α_b =0.1 and α_d =0.65), LAI is the leaf area index, $G = \varphi_1 + \varphi_2 cos\theta$ is the leaf orientation factor depending on solar zenith angle (Dai et al. 2004), $\varphi_1 = 0.5 - 0.633\chi - 0.33\chi^2$, $\varphi_2 = 0.877(1 - 2\varphi_1)$, and χ is an empirical leaf angle distribution parameter ranging from -1 to 1 (-1 for vertical distributed leaves, 0 for spherical leaf angle distribution with randomly distributed leaves, and 1 for horizontal distributed leaves), $k_b = G/cos\theta$ is the extinction coefficient of black leaves, and $E_i(n,x)$ is exponential integral with n a nonnegative integer (Nijssen and Lettenmaier 1999), defined as:

$$E_i(n,x) = \int_1^\infty \frac{e^{-xt}}{t^n} dt \qquad \text{(Eq. 2.18)}$$

$\tau' = \tau_b'$ is used in eq. 2.13 and eq. 2.14 when calculating the transmission and reflection factors of direct radiation (τ_b and β_b), while $\tau' = \tau_d'$ is used for diffuse radiation. Since the approach for diffuse radiation is just an integral of single beam components over the hemisphere, so the reflection factors for an infinitely deep canopy β_b' and β_d' are estimated using the same equation (eq. 2.17).

Then the transmitted and reflected radiation are calculated by multiplying the corresponding factors to the incoming direct or diffuse radiation, respectively.

Two-leaf model

To estimate the radiation absorbed by the canopy, we applied a two-leaf model based on Wang and Leuning (1988) that separates the canopy into two groups of leaves including sunlit leaves and shaded leaves, which receive different components and portions of incoming shortwave radiation. It is assumed that the sunlit leaves receive both direct and diffuse solar radiation, while shaded leaves absorb the diffuse radiation only (Spitters 1986).

The leaf area index (LAI) of sunlit and shaded leaves of the canopy were derived by Dai et al. 2004:

$$LAI_{sun} = \frac{1}{K_b}(1 - e^{-K_b LAI}) \tag{Eq. 2.19}$$

$$LAI_{shade} = LAI - LAI_{sun} \tag{Eq. 2.20}$$

Then the total solar radiation flux density absorbed by the sunlit leaves in the canopy is given as the sum of the direct component of direct radiation $AS_{b,b}$, the scattered component of direct radiation $AS_{b,s}$, and the diffuse radiation AS_d, which is:

$$AS_{sun} = AS_{b,b} + AS_{b,s} + AS_d \tag{Eq. 2.21}$$

And the solar radiation flux absorbed by the shaded leaves in the canopy is given as the sum of the scattered component of direct radiation $AS_{b,s}$ and the diffuse radiation AS_d, represented as:

$$AS_{shade} = AS_{b,s} + AS_d \tag{Eq. 2.22}$$

The absorption of the diffuse radiation and the scattered component of direct radiation is averaged over the total leaf area, while the absorption of the direct component of direct radiation is given per unit of sunlit leaf area only.

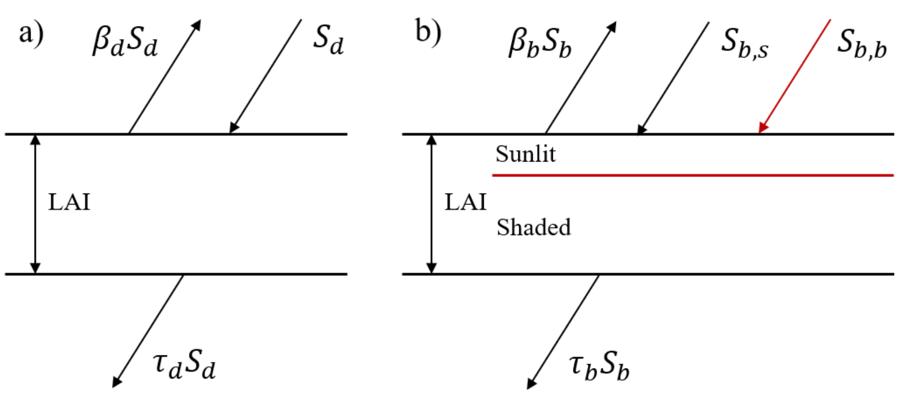


Figure 2.5 Illustration of the radiation transfer through a canopy for diffuse (a) and direct (b) radiation.

We coupled the canopy radiative transfer model into the two-leaf model to calculate the absorption of different radiation components by the sunlit and shaded leaves (Figure 2.5).

Absorption is the complement to transmission and reflection, hence the absorbed diffuse radiation AS_d is given by:

$$AS_d = S_d(1 - \tau_d - \beta_d) \quad \text{(Eq. 2.23)}$$

The absorbed direct component (excluding scattering) of direct radiation $AS_{b,b}$ can be expressed as (Spitters 1986 (second part in eq.14)):

$$AS_{b,b} = S_b K_b \quad \text{(Eq. 2.24)}$$

And the absorbed scattered component of direct radiation $AS_{b,s}$ can be calculated as:

$$AS_{b,s} = S_b(1 - \tau_b - \beta_b) - AS_{b,b} \quad \text{(Eq. 2.25)}$$

2.2.3. Stomatal conductance

In our study, the Ball-Berry-Leuning (BBL) stomatal conductance model was coupled in the photosynthesis process (Leuning 1995), and the stomatal conductance is given by:

$$g_s = g_0 + g_1 \cdot \frac{A_n}{(1+\frac{VPD}{D_0})\cdot(C_a - \Gamma^*)} \quad \text{(Eq. 2.26)}$$

where A_n is the net leaf CO_2 assimilation rate, VPD is vapor pressure deficit, C_a is CO_2 concentration at the leaf surface, Γ^* is the CO_2 compensation point, and g_0, g_1 and D_0 are empirical coefficients and their values (Panek and Goldstein 2001) are presented in Table 2.3. We also provide Ball-Berry model (Ball 1988) as another option for calculating stomatal conductance. In Ball-Berry model, the stomatal conductance is given by:

$$g_s = g_0 + g_1 \cdot \frac{A_n \cdot RH}{C_a} \quad \text{(Eq. 2.27)}$$

where RH is relative humidity, and other parameters are the same as in the BBL model.

Table 2.3 Parameters for stomatal conductance and photosynthesis.

Parameter	Value	Unit	References
g_0	7.5	--	Panek and Goldstein 2001
g_1	0.01	mol m^{-2} s^{-1}	Panek and Goldstein 2001
D_0	2	KPa	Panek and Goldstein 2001
θ	0.7	--	Medlyn et al. 2002
α	0.3	mol mol^{-1}	Medlyn et al. 2002
k_p	0.7	--	Oleson et al. 2013
O_i	210	mmol mol^{-1}	--
R_{gas}	8.314	J K^{-1} mol^{-1}	--
K_{o25}	248	mmol mol^{-1}	Thornton 2010
K_{c25}	404	µmol mol^{-1}	Thornton 2010
Q_{10,K_o}	1.2	--	Thornton 2010
Q_{10,K_c}	2.1	--	Thornton 2010
Q_{10,R_d}	2.0	--	Thornton 2010
Q_{10}	2.0	--	Oleson et al. 2013
S_1 for V_{cmax}	0.3	K^{-1}	Oleson et al. 2013
S_2 for V_{cmax}	313.15	K	Oleson et al. 2013
S_3	0.2	K^{-1}	Oleson et al. 2013
S_4	288.15	K	Oleson et al. 2013
S_1 for R_d	1.3	K^{-1}	Oleson et al. 2013
S_2 for R_d	328.15	K	Oleson et al. 2013

2.2.4. Photosynthesis

The photosynthesis model was performed for the C3 and C4 vegetation separately. Then, the C3 and C4 distribution map was used to get the relative proportions of C3 and C4 in each pixel, and the sum of the results provided the final photosynthesis carboxylation at each pixel.

We used the biochemical photosynthesis model for C3 plants based on the model of Farquhar et al. (1980) and C4 plants based on the model of Collatz et al. (1992). The net leaf photosynthesis A_n could be modeled as the minimum of three limiting rates after accounting for dark respiration (R_d, leaf daytime maintenance respiration):

$$A_n = min(A_c, A_j, A_p) - R_d \qquad \text{(Eq. 2.28)}$$

A_c is the rate of photosynthesis when the RuBP carboxylase (Rubisco) is limited, which is given by:

$$A_c = V_{cmax} \cdot \frac{C_i - \Gamma^*}{C_i + K_c(1 + O_i/K_o)} \quad \text{for C3 plants} \qquad \text{(Eq. 2.29a)}$$

$$A_c = V_{cmax} \quad \text{for C4 plants} \qquad \text{(Eq. 2.29b)}$$

where V_{cmax} is the maximum rate of carboxylation, C_i is the intercellular CO_2 concentration, O_i is the atmospheric concentration of O_2, Γ^* is the CO_2 compensation point in the absence of dark respiration, and K_c and K_o are the Michaelis–Menten constants for rubisco carboxylation and oxygenation, respectively, scaled by the temperature using a Q_{10} relationship. The Rubisco activity K_c and $K_o$can be calculated following the Michaelis–Menten dynamics for CO_2 and O_2, respectively (Thornton 2010). The calculation of K_c varies depending on the temperature threshold of 15 degree C. The equations used to calculate K_c and K_o are described:

$$K_o = K_{o25} \times Q_{10,K_o}{}^{\frac{T_c - 25}{10}} \qquad \text{(Eq. 2.30)}$$

$$K_c = \begin{cases} K_{c25} \times Q_{10,K_c}^{\frac{T_c-25}{10}} & for\ T_c > 15℃ \\ \frac{K_{c25} \times (1.8 \times Q_{10,K_c})^{\frac{T_c-15}{10}}}{Q_{10,K_c}} & for\ T_c \leq 15℃ \end{cases} \qquad \text{(Eq. 2.31)}$$

where T_c is the leaf temperature in Celsius degrees.

A_j is the rate of photosynthesis when the regeneration of RuBP is limited (light-limited), which is given by:

$$A_j = J \cdot \frac{C_i - \Gamma^*}{4C_i + 8\Gamma^*} \quad \text{for C3 plants} \qquad \text{(Eq. 2.32a)}$$

$$A_j = 0.067 * Q \quad \text{for C4 plants} \qquad \text{(Eq. 2.32b)}$$

where J is the rate of electron transport, and it depends on the photosynthetically active radiation absorbed by the leaf expressed as (Medlyn et al. 2002):

$$\theta J^2 - (\alpha Q + J_{max}) + \alpha Q J_{max} = 0 \qquad \text{(Eq. 2.33)}$$

where J_{max} is the maximum potential rate of electron transport, Q is the photosynthetically active photon flux density, θ is the curvature parameter of the light response curve, and α is the quantum yield of electron transport.

A_p is the rate of photosynthesis when the product is limited for C3 plants and when the PEP carboxylase is limited for C4 plants, which is given by:

$$A_p = 0.5V_{cmax} \quad \text{for C3 plants} \qquad \text{(Eq. 2.34a)}$$

$$A_p = k_p \times 10^6 \times \frac{C_i}{P_{atm}} \quad \text{for C4 plants} \qquad \text{(Eq. 2.34b)}$$

where k_p is the initial slope of C4 CO_2 response curve and P_{atm} is the atmospheric pressure. The values of the parameters are listed in Table 2.3.

We also coupled the Eq. 2.37 that represents the CO_2 diffusion constraints of photosynthetic rate:

$$A_{(c\ or\ j)} = (C_a - C_i) \times G_{co2} \qquad \text{(Eq. 2.35)}$$

where, C_a is the atmospheric CO_2 concentration and G_{co2} is the velocity of CO_2 diffusion from atmosphere into leaves.

We used a temperature function and high temperature stress function to scale and describe the temperature dependences of V_{cmax}, J_{max}, Γ^*, and R_d (Bernacchi et al. 2001, Bonan et al. 2011, Medlyn et al. 2002, Oleson et al. 2013). For C3 plants, the equations are expressed as below:

$$V_{cmax} = V_{cmax25} \times f(T) \times f_H(T) \quad \text{(Eq. 2.36)}$$

$$J_{max} = J_{max25} \times f(T) \times f_H(T) \quad \text{(Eq. 2.37)}$$

$$R_d = R_{d25} \times f(T) \times f_H(T) \quad \text{(Eq. 2.38)}$$

$$\Gamma^* = \Gamma^*{}_{25} \times f(T) \quad \text{(Eq. 2.39)}$$

with the temperature functions described as:

$$f(T) = e^{\frac{\Delta H_a \times (T_k - 298.15)}{298.15 \times R_{gas} \times T_k}} \quad \text{(Eq. 2.40)}$$

$$f_H(T) = \frac{1 + e^{\frac{298.15 \times \Delta S - \Delta H_d}{298.15 \times R_{gas}}}}{1 + e^{\frac{\Delta S \times T_k - \Delta H_d}{R_{gas} \times T_k}}} \quad \text{(Eq. 2.41)}$$

Where T_k is the leaf temperature in Kelvin, R_{gas} is the universal gas constant, the values of temperature dependence parameters ΔH_a, ΔH_d, and ΔS are listed in Table 2.4, and the J_{max25}, R_{d25}, and $\Gamma^*{}_{25}$ are the corresponding parameters at 25 degree C and are calculated as: $J_{max25} = 1.97 V_{cmax25}$, $R_{d25} = 0.015 V_{cmax25}$, and $\Gamma^*{}_{25} = 42.75$.

Table 2.4 Temperature dependence parameters.

Parameter	ΔH_a (J/mol)	ΔH_d (J/mol)	ΔS (J/mol/K)
V_{cmax}	65330	149250	485
J_{max}	43540	152040	495
R_d	46390	150650	490
Γ^*	37830	-	-

For C4 plants, the temperature dependence of V_{cmax} is scaled by high temperature stress function and low temperature stress function,

$$V_{cmax} = V_{cmax25}\left[\frac{Q_{10}^{\frac{T_k-298.15}{10}}}{f_H(T)\times f_L(T)}\right] \quad \text{(Eq. 2.42)}$$

$$f_H(T) = 1 + e^{s_1\times(T_k-s_2)} \quad \text{(Eq. 2.43)}$$

$$f_L(T) = 1 + e^{s_3\times(s_4-T_k)} \quad \text{(Eq. 2.44)}$$

where the values of s_1, s_2, s_3, and s_4 are listed in Table 2.3, and the temperature dependence of dark respiration is expressed as:

$$R_d = R_{d25}\left[\frac{Q_{10}^{\frac{T_k-298.15}{10}}}{f_H(T)}\right] \quad \text{(Eq. 2.45)}$$

The maximum rate of carboxylation at 25 degree C (V_{cmax25}) depends on the leaf nitrogen concentration and specific leaf area,

$$V_{cmax25} = N_a \times F_{LNR} \times F_{NR} \times a_{R25} \quad \text{(Eq. 2.46)}$$

where N_a is the leaf nitrogen concentration (gN per m^2 leaf area), F_{LNR} is the fraction of leaf nitrogen in Rubisco (gN in Rubisco per gN in leaf), $F_{NR} = 7.16$ is the weight proportion of Rubisco to its nitrogen content (g Rubisco per gN in Rubisco), and $a_{R25} = 60$ is the specific activity if Rubisco (μmol CO_2 per g Rubisco per second). The leaf nitrogen concentration N_a is a function of C/N ration and specific leaf area:

$$N_a = \frac{1}{CN_l \times SLA} \quad \text{(Eq. 2.47)}$$

the CN_l is the ration of carbon to nitrogen in the leaf (gC/gN) and SLA is the specific leaf area (m^2 leaf area per gC). The CN_l, SLA, and F_{LNR} varies for different vegetation type, and the values can be found in Table 2.5 (Thornton 2010, Oleson et al. 2013).

Table 2.5 Photosynthetic parameters for V_{cmax25}.

Vegetation type	CN_l	SLA	F_{LNR}
ENF	42	0.012	0.040
EBF	35	0.012	0.046
DNF	25	0.024	0.055
DBF	24	0.030	0.080
MF	32	0.020	0.060
CSH	42	0.012	0.040
OSH	42	0.012	0.040
WSA	25	0.030	0.090
SAV	25	0.030	0.090
GRA	24	0.045	0.120
WET	42	0.012	0.040
CRO	25	0.070	0.410

2.3. Statistical analysis

The Pearson correlation coefficient. The Pearson correlation coefficient (r) is commonly used to measure the linear correlation between two sets of data. Here, we used Pearson's r to evaluate the climate variables from GLDAS 2.1 datasets to the observations from flux tower sites, and it could be expressed as:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \quad \text{(Eq. 2.48)}$$

where x_i and y_i denote the ith data in the two datasets, respectively, $\bar{x}$ and $\bar{y}$ represent the mean of the two datasets, respectively.

The root mean square error & The Bias. The root mean square error (RMSE) and the Bias are frequently used to measure the differences between estimated data and observed data. We used RMSE and Bias in our research to measure the differences between the climate variables derived from GLDAS 2.1 or our estimated LAI and GPP to the flux tower observations. Their equations were:

$$RMSE = \sqrt{\frac{\sum(x_i - \hat{x}_i)^2}{N}} \quad \text{(Eq. 2.49)}$$

$$Bias = \frac{\sum(\hat{x}_i - x_i)}{N} \quad \text{(Eq. 2.50)}$$

where x_i and $\hat{x}_i$ are the ith observed data and estimated data, respectively, and N is the number of the samples. And the relative RMSE and the relative Bias were obtained by dividing the means of the variables.

The goodness of fit. The goodness of fit (R^2) is an important index that describes how well a model fits a set of observations, and measuring the goodness of fit could summarize the discrepancy between observed data and model predicted data. In this study, the goodness of fit was used to assess the performance of the double logistic fitting to LAI, and was calculated as:

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \quad \text{(Eq. 2.51)}$$

where y_i denotes the ith observation data, $\hat{y}_i$ represents the ith data predicted by the double logistic function, and $\overline{y}$ is the mean of the observation samples.

The trends of the variables (climate variables and GPP) in our research were determined by linear regression analyses and corresponding F tests were used to test the statistical significance of the trends at a 1% significance level.

2.4. Sensitivity analysis

We performed a simple sensitivity analysis for global GPP estimated by our model in 2013 by using the method provided by Ryu et al. 2011. Six variables including air temperature, solar radiation, atmospheric CO_2 concentration, vapor pressure deficit (VPD), leaf area index (LAI), and V_{cmax25} were selected to examine the sensitivities of our model to these key environmental and biophysiological drivers. We changed the values of each variable by $\pm$30% while keeping other variables the same, and compared the GPP outputs.

3. Investigating the variation of climate variables from GLDAS 2.1 and assessing the performance of fitting double logistic functions of LAI

Key model input data, including GLDAS 2.1-derived reanalysis climate data and satellite-based MODIS LAI data, were evaluated and improved to reduce the biases before inputting to the process-based ecosystem model. This chapter evaluated five GLDAS 2.1-derived climate variables, including air temperature, precipitation, downward shortwave radiation, air pressure, and VPD (derived from specific humidity), against the observations from 120 worldwide flux tower sites. The temporal and spatial variations of the five climate variables from 2001 to 2020 were also investigated at global scale. Moreover, remote sensing LAI product (MCD15A3H) was reconstructed to obtain a high-quality and continuous time series by fitting double logistic functions after eliminating noise and outliers.

3.1. Evaluating the climate variables from GLDAS 2.1 using flux tower observations

The air temperature, precipitation, downward shortwave radiation, air pressure, and VPD (derived from specific humidity) obtained from GLDAS 2.1 were evaluated by the flux tower observations. We integrated the 3-hour GLDAS 2.1 data to daily scale and extracted the $0.25^{o} \times 0.25^{o}$ pixels using the geographic coordinates of the 120 selected flux tower sites, then we compared the GLDAS estimates with flux tower observations over daily scale. The selected flux towers cover 13 major vegetation types including evergreen needleleaf forest (ENF), evergreen broadleaf forest (EBF), deciduous needleleaf forest (DNF), deciduous broadleaf forest (DBF), mixed forest (MF), grassland (GRA), cropland (CRO), closed shrubland (CSH), open shrubland (OSH), wetland (WET), savanna (SAV), woody savanna (WSA), and snow or ice covered land (SNO). Detailed information for the sites is shown in Table A.1.

The GLDAS-derived air temperature was highly correlated to the flux tower observations, with the Pearson's correlation coefficient larger than 0.85 for all the flux sites (Figure 3.1). The RMSE ranged from 0.85°C to 5.30°C with the mean of 2.10°C. And the bias varied from -3.98°C to 2.07°C with the mean of -0.20°C.

The GLDAS-derived solar radiation also showed a strong relationship to the flux tower observations, with an average r of 0.91 (Figure 3.2). The RMSE varied from 23.31W/m^2 to 75.16W/m^2 with the mean of 37.96W/m^2. And the bias ranged from -22.92 W/m^2 to 18.02 W/m^2 with the mean of -0.66 W/m^2.

For the air pressure, the average of the correlation coefficient between GLDAS-derived data and flux tower observations was 0.91, but with relative large variation (Figure 3.3). The RMSE varied from 0.04KPa to 3.22KPa with an average of 0.74KPa. And the bias ranged from -2.91 KPa to 3.20KPa with the mean of -0.01KPa.

For the VPD, we found good relationships between GLDAS data and flux tower data on the sites of all the vegetation types, except NO-Blv where it was mainly covered by snow or ice (Figure 3.4). The average of Pearson’s r over all the sites was 0.86. The RMSE varies from 0.07KPa to 0.53KPa with an average of 0.24KPa. And the bias ranged from -0.32KPa to 0.25KPa with the mean of -0.04KPa.

For the precipitation, the GLDAS-derived data showed a poor relationship to the flux tower observations for all vegetation types at daily scale, with an average r of 0.44 (Figure 3.5). The RMSE varies from 1.81mm to 13.08mm with an average of 5.35mm,and the bias ranges from -0.86mm to 1.44mm with the mean of 0.43mm. The GLDAS 2.1 data exhibited an overall overestimation on precipitaion with 85% sites showing positive bias.

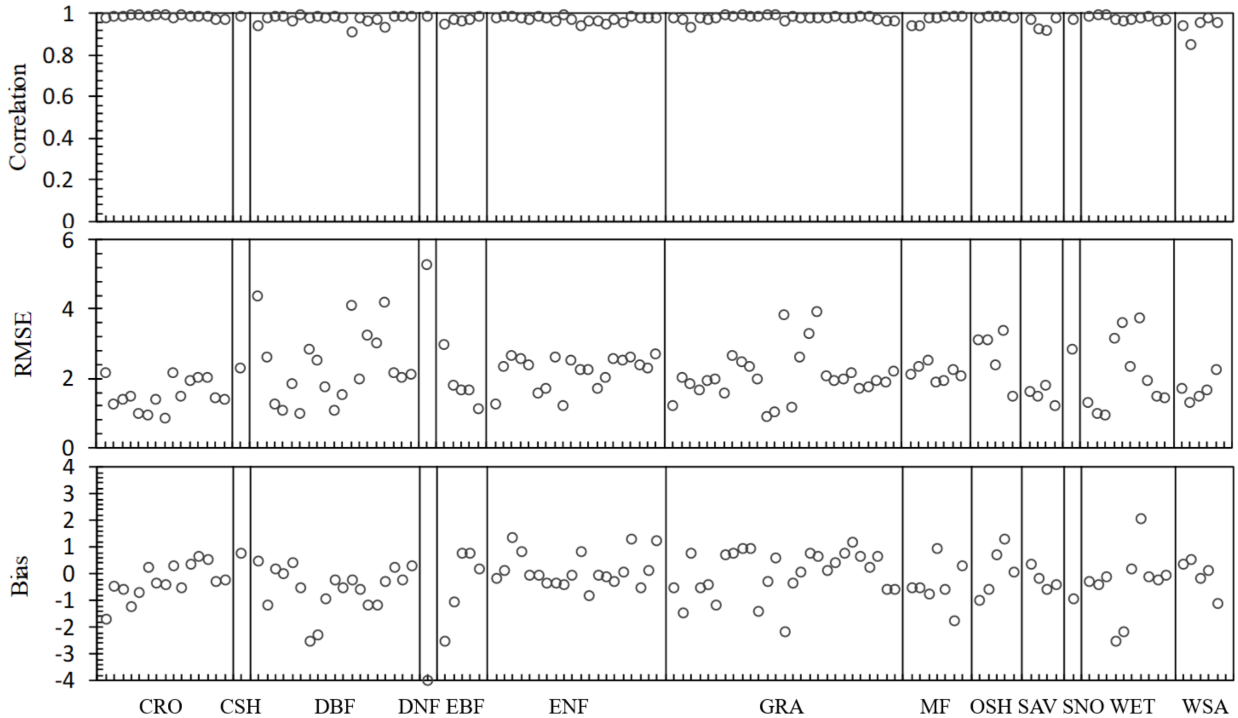


Figure 3.1 Comparison of daily air temperature (°C) between the observations from flux towers and the data retrieved from GLDAS 2.1.

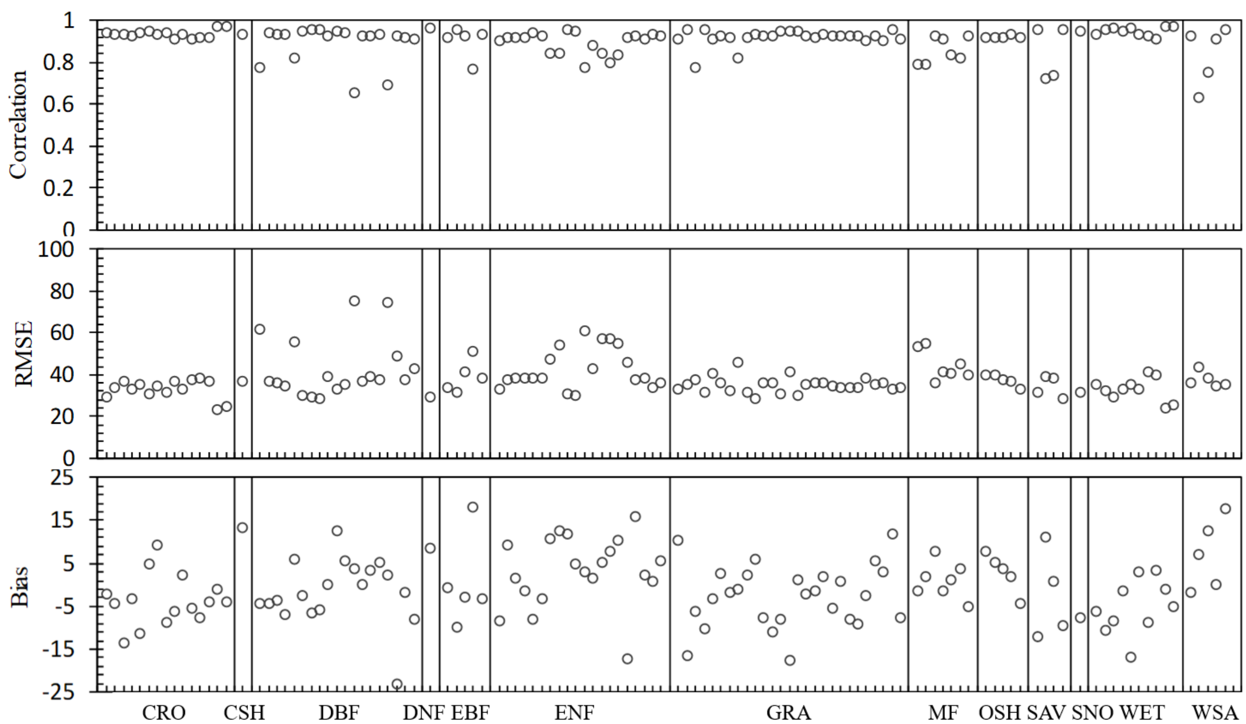


Figure 3.2 Comparison of daily downward shortwave radiation (W/m^2) between the observations from flux towers and the data retrieved from GLDAS 2.1.

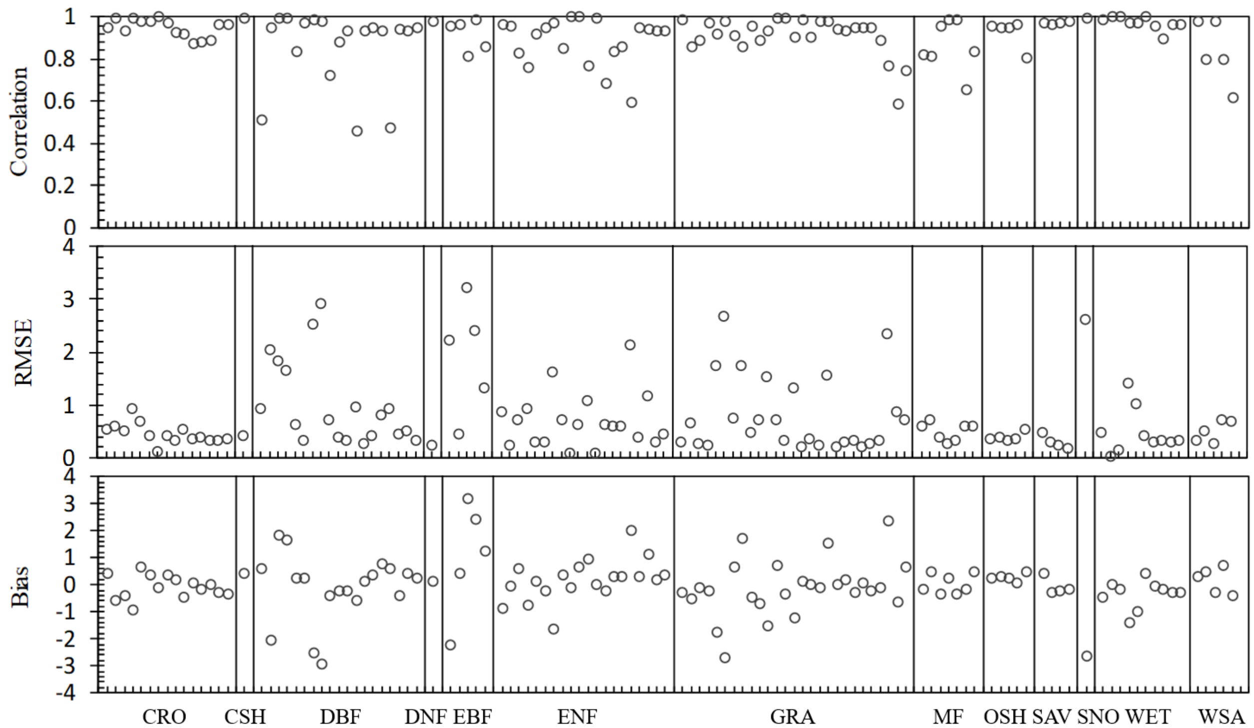


Figure 3.3 Comparison of daily air pressure (KPa) between the observations from flux towers and the data retrieved from GLDAS 2.1.

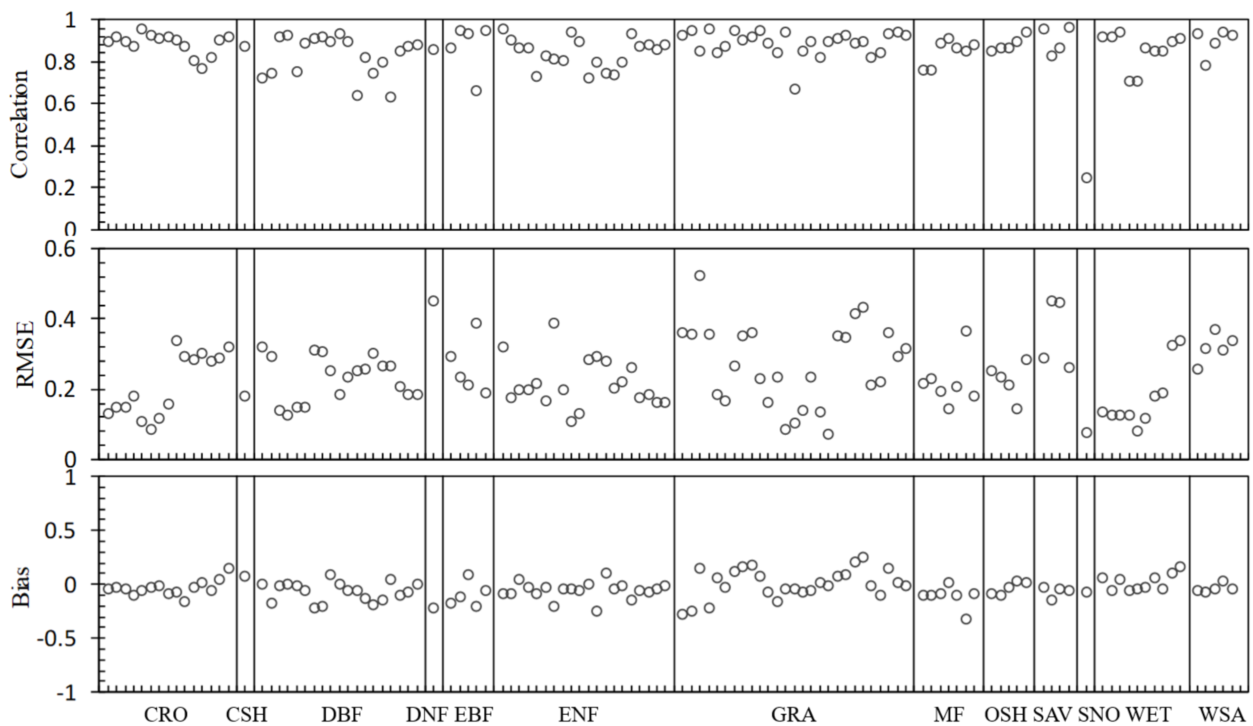


Figure 3.4 Comparison of daily VPD (KPa) between the observations from flux towers and the data retrieved from GLDAS 2.1.

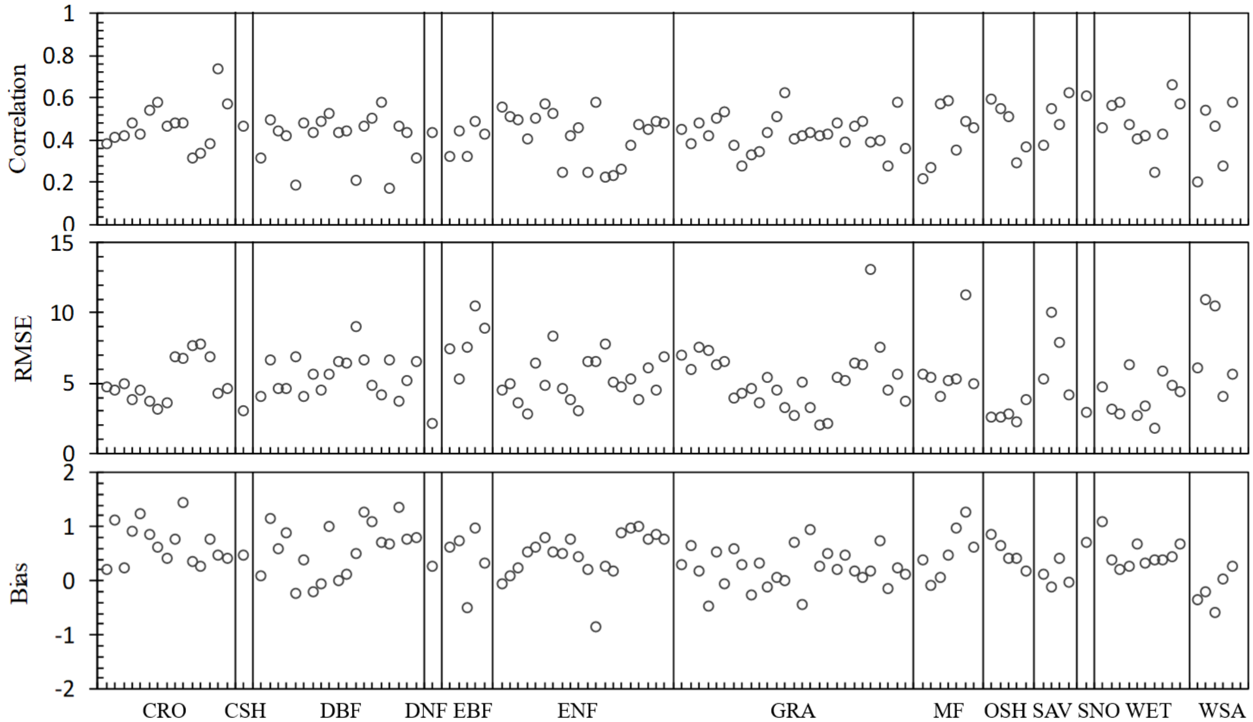


Figure 3.5 Comparison of daily precipitation (mm) between the observations from flux towers and the data retrieved from GLDAS 2.1.

Though there exist some variation between the two datasets, the five climate variables derived from GLDAS showed strong linear relationship with the flux tower observations at annual scale and the relationships were all statistically significant ($P<0.01$) (Figure 3.6). The R^2 was larger than 0.93 for all the variables except precipitation ($R^2 = 0.77$). Notably, we found that the relative RMSE was high in precipitation (32.63%) and precipitation showed large and positive relative bias (22.92%) (Table 3.1), revealing that the GLDAS 2.1 tended to overestimate the precipitation with an average bias of 222.22mm. The discrepancy might be caused by the high spatial variability of precipitation, which made it difficult to quantify in large scale. Besides the annual total precipitation amount, the GLDAS 2.1-derived annual air temperature, solar radiation, air pressure, and VPD were able to provide highly correlated results but seemed to be slightly underestimated with an average bias of -0.20°C, -0.66W/m^2, -0.01KPa, and -0,04KPa, respectively.

Table 3.1 The RMSE and bias of the five climate variables between GLDAS 2.1 and flux tower sites.

Variables (Units)	RMSE	Relative (%)	Bias	Relative (%)
Temperature (°C)	0.94	10.18	-0.20	-2.12
Radiation (W/m^2)	7.79	4.88	-0.66	-0.42
Pressure (Kpa)	0.97	1.00	-0.01	-0.01
VPD (Kpa)	0.11	18.66	-0.04	-6.33
Precipitation (mm)	222.22	32.63	156.12	22.92

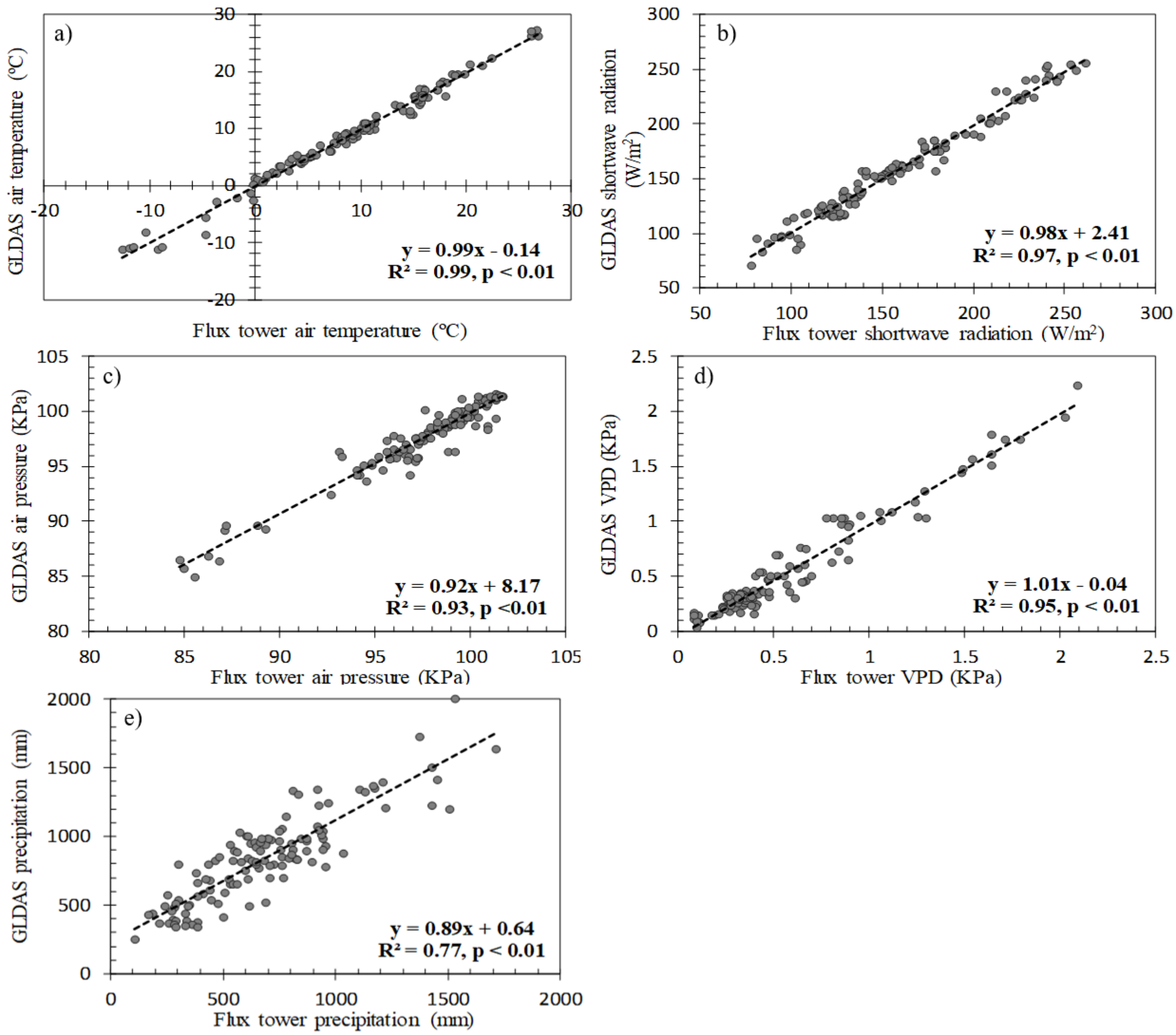


Figure 3.6 Comparison of annual means of a) air temperature, b) downward shortwave radiation, c) air pressure, d) vapor pressure deficit (VPD), and e) precipitation amount between the observations from flux towers and the data retrieved from GLDAS 2.1.

Our results indicated that the GLDAS 2.1 data had a good agreement with the in-situ flux tower observations on air temperature, solar radiation, air pressure, and VPD at both daily and annual scales, while it failed to accurately capture the precipitation amount. The obvious overestimation in precipitation data should be carefully considered before using.

3.2. Characterizing the temporal and spatial variations of climate variables from GLDAS 2.1

The GLDAS 2.1-derived global annual mean air temperature varied from 12.87°C to 14.37°C, with an average of 13.59°C, during the study period from 2001 to 2020. The air temperature exhibited a significant warming trend ($p < 0.01$) with an increase of 11% during this period (Figure 3.7a). We also found that the increase rate of the second decade (0.6°C per decade from 2011 to 2020) doubled the increase rate of the first decade (1.2°C per decade from 2000 to 2010), indicating a continuous warming in recent decades.

The global annual downward shortwave radiation, known as solar radiation, ranged from 185.98W/m^2 to 195.06W/m^2, with an average of 189.84W/m^2 within the 20 years. Interestingly, we found a jump in global annual solar radiation after 2010. The mean of the solar radiation after 2010 was 192.33W/m^2, which was 2.66% higher than the mean of the solar radiation before 2010. Although the global annual solar radiation significantly ($p < 0.01$) increased 3% during 2001-2020 (Figure 3.7b), the solar radiation did not show a significant trend ($p = 0.23$) in the first decade while it exhibited a significant dimming trend ($p = 0.01$) of -0.42W/m^2/yr during the second decade. Our result was in agreement with the study from Yuan et al. (2021), who found a brightening trend since 1982 till 2019. However, the solar radiation in the 2000s did not have too much differences compared to that in the 2010s in Yuan et al.'s paper. This discrepancy might be from the different data sources of the studies.

The global air pressure did not show a significant change ($p = 0.47$) during the study period from 2001 to 2020, and it varies from 93.96KPa to 94.06KPa with an average of 94.01KPa (Figure 3.7c). However, the global annual air pressure exhibited a decreasing trend by 0.05KPa/decade in the first decade and then increased by 0.08KPa/decade in the second decade, and both of the changes were statistically significant at 5% significance level with $p = 0.02$ for both.

The global annual averaged vapor pressure deficit varies from 0.65KPa to 0.96KPa with an average of 0.80KPa during the period of 2001-2020. The annual VPD shows significantly increasing trend ($p < 0.01$) in a rate of 0.15KPa/decade from 2001 to 2020, and the increasing trend was consistent during the two decades (Figure 3.7d). Our study confirmed the findings in Yuan et al. (2019) that the global VPD strongly increased after the late 1990s.

The global annual precipitation amount varied from 816.92mm to 991.28mm, with an average of 870.91mm, during the study period from 2001 to 2020. The precipitation showed a significant increasing trend ($p = 0.013$) at 5% significance level during this period, and the trend increased by 49.57mm/decade (Figure 3.7e). However, the changes in both of the two decades during our study period were not statistically significant. The precipitation was relatively stable in the first 15 years, and rapidly increased 14% after 2015 followed by a 14% drop in 2020.

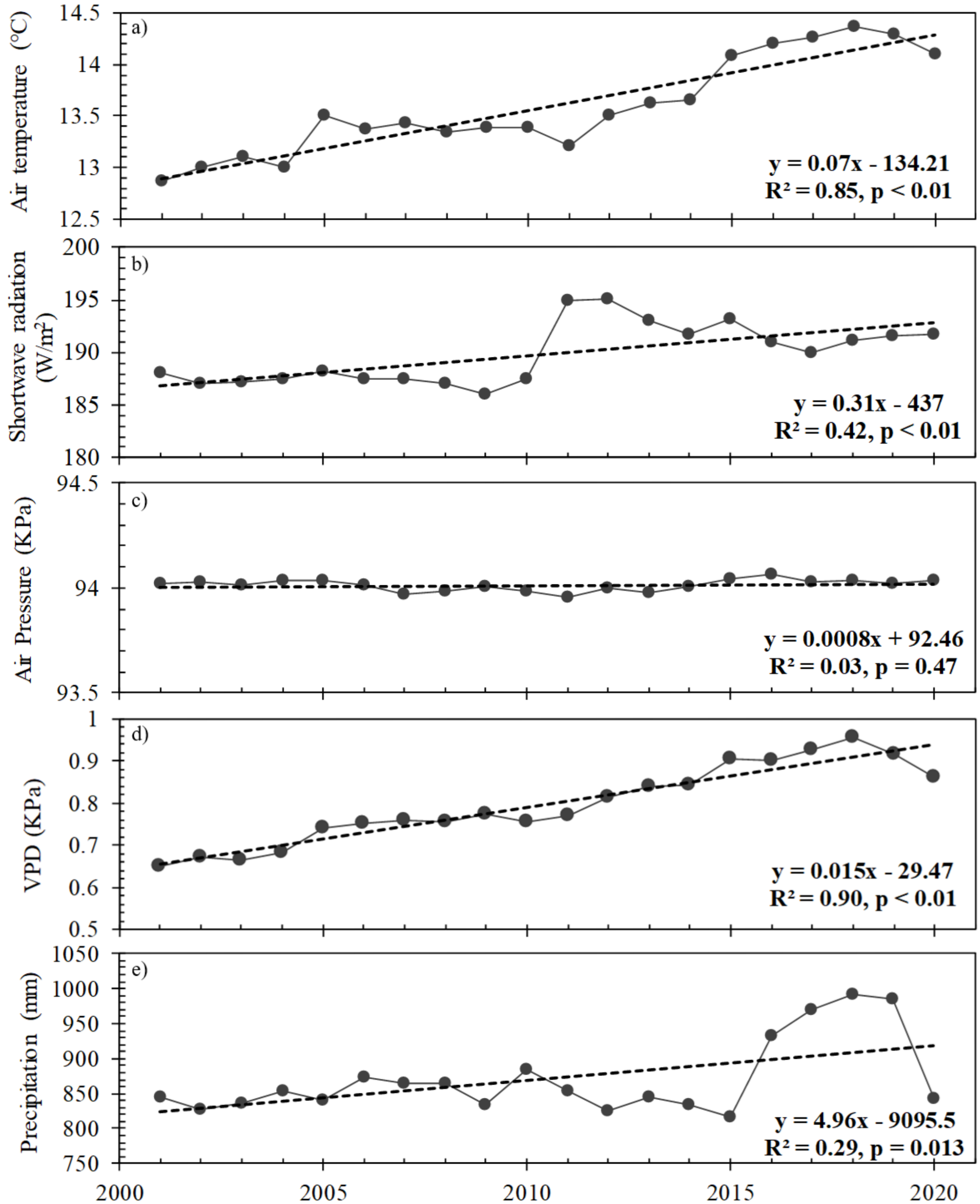


Figure 3.7 Time series of global annual a) air temperature, b) downward shortwave radiation, c) air pressure, d) vapor pressure deficit (VPD), and e) precipitation amount retrieved from GLDAS 2.1 from 2001 to 2020. The dashed lines indicate the annual trends for the corresponding climate variables.

The five climate variables showed different spatial variations in global scale and their changes over the past decades were not homogenous in space. The annual average air temperature ranged from -26.09°C to 34.16°C during 2001 to 2020. The annual temperature showed an apparent decreasing gradient from the equator along the latitude with some exceptions at high elevations such as the Tibetan plateau in China, the Andes Mountains in the western South America, and the North American Cordillera in the western North America (Figure 3.8-a1). During the recent two decades, the air temperature showed a significant warming trend in the western North America, most regions in South America, Africa, most of Europe, Middle East, the southern Asia, and Australia, while only a small area in southern Canada and eastern Brazil exhibited a significant cooling trend (Figure 3.8-a2). Interestingly, we found that the most pronounced warming trend appeared at the high elevation regions, including Tibetan plateau, Andes Mountains, Iranian plateau and Arabian plateau. This finding confirmed the previous studies that the rate of global warming was amplified with elevation (Rangwala and Miller 2012, Pepin et al. 2015, Palazzi et al. 2019), and this elevation-dependent warming still existed in recent decades. The annual average solar radiation ranged from 56.30W/m^2 to 305.82W/m^2 during the study period of 2001 to 2020. The gradients in solar radiation also occurred along latitudes, and was affected by aerosols and clouds in the atmosphere and the geographic elevations (Figure 3.8-b1). Solar radiation significantly increased in the regions near 70°N in latitude, most of Africa, northern South America, southern India, and western Australia, while eastern China, some regions in Europe, and the middle north of America showed significantly dimming trend during the twenty years (Figure 3.8-b2). It was interesting to find that the regions with significantly decreasing solar radiation were also the most aerosols affected areas in recent decades (Subba et al. 2020), indicating the important role of aerosols in characterizing the global solar radiation variations.

The global averaged air pressure over 2001 to 2020 varied from 47.61KPa to 102.87KPa. The global pattern of air pressure was strongly dependent on the geographic elevation, with mountains and plateaus showing lower air pressure and flat terrains and plains showing higher air pressure (Figure 3.8-c1). The significant changes in air pressure were scattered distributed in the world, and most regions did not show a significant change (Figure 3.8-c2). However, we found that the high elevation regions showed relatively distinct significant trends but the trends were not uniform within each region.

During the period of 2001 to 2020, the global averaged vapor pressure deficit ranged from 0 to 3.64KPa. Figure 3.8-d1 showed the global distribution pattern of VPD, and the VPD was highest in the arid areas, such as the Sahara Desert, the Arabian Desert, and the Deserts of Australia. The trends of VPD during the two decades showed significant increasing trend in most tropical and subtropical regions with up to 0.94KPa per decade increasing rate (Figure 3.8-d2). The regions in the Brazilian Highlands in Brazil exhibited significant negative trend.

The global annual precipitation amount averaged over 2001-2022 varied from 1.47mm to 8721.65mm. The regions with the most abundant rainfall distributed near the equator, including the Amazon Rainforest, the Malay Archipelago, and the Congo Basin in Africa, while the areas with least rainfall located at arid desert regions (Figure 3.8-e1). In addition, changes in annual total precipitation amount did not have a centralized significant changes spatially (Figure 3.8-e2). The significant wetting trend scattered in the western North America, the eastern Asia, the East African Plateau, and the Brazilian Highlands, while a significant drying trend occurred in the Amazon Rainforest, the Congo Basin in Africa, and the southern South America.

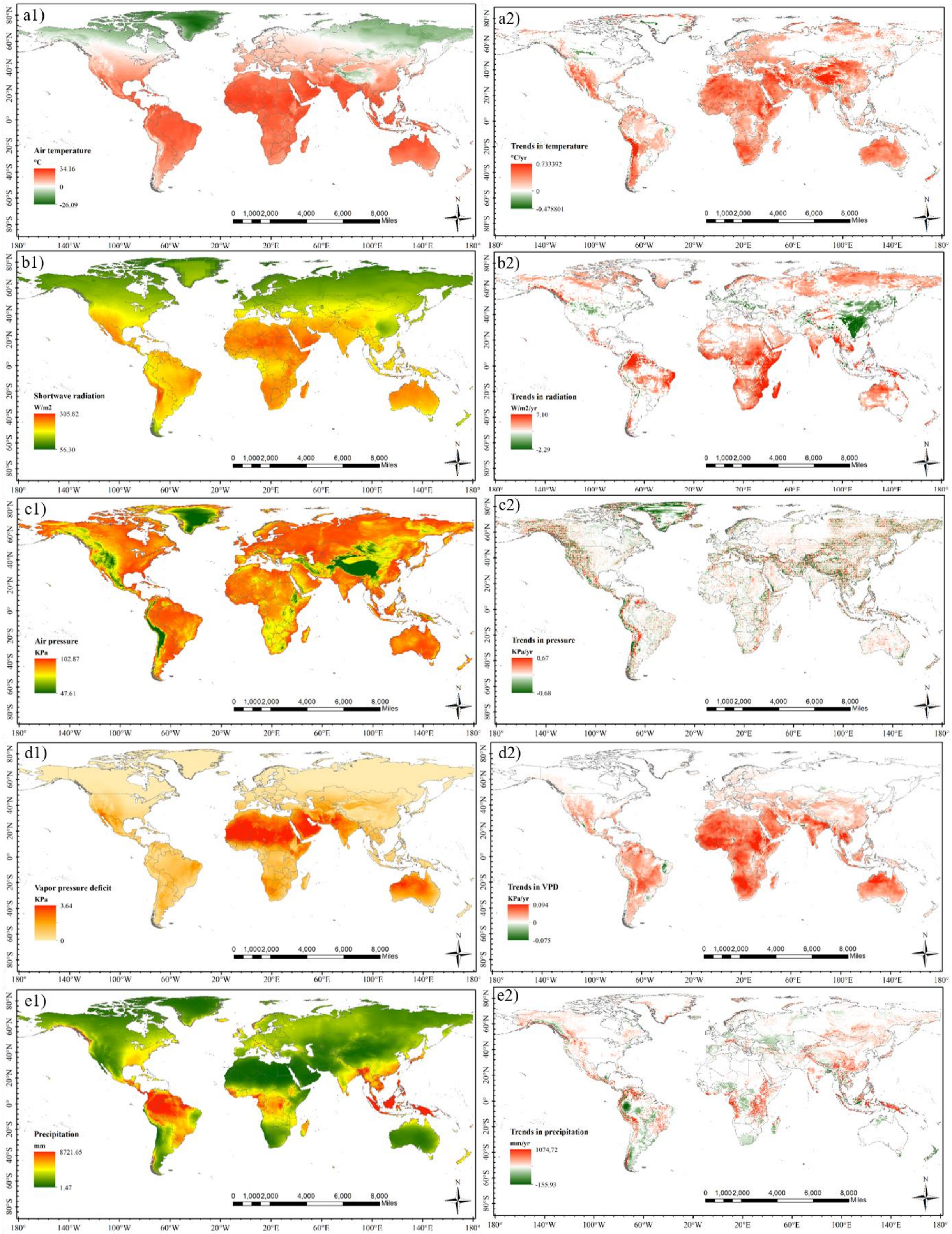


Figure 3.8 Spatial distributions of global annual average of a1) air temperature, b1) downward shortwave radiation, c1) air pressure, d1) VPD, and e1) precipitation and global annual trends (a2 - e2) during the period of 2001-2020. For the global annual trend maps, the colors showed the regions where trends were

statistical significant at 5% significant level, while the trends in the regions with white color were not significant ($p > 0.05$).

3.3. Assessing the performance of reconstructing LAI

3.3.1. Single vegetation cycle

For the grids with a single valid vegetation cycle, we selected 16 typical regions randomly, which covered major climate and vegetation types, and evaluated the performance of LAI reconstructing in 2003. The sample sites covers 8 major vegetation types, including evergreen broadleaf forest, evergreen needleleaf forest, deciduous broadleaf forest, savanna, woody savanna, grassland, cropland, and wetland, and the detailed information of the selected regions are shown in Table 3.2.

Table 3.2 Detailed information of the selected sample sites with single vegetation cycle.

Site ID	Longitude	Latitude	Vegetation type
1	131.125	-13.375	Grassland
2	117.651	40.982	Deciduous broadleaf forest
3	107.547	14.573	Evergreen broadleaf forest
4	105.917	39.147	Wetland
5	105.580	31.444	Cropland
6	100.958	70.082	Woody savanna
7	62.092	61.674	Evergreen needleleaf forest
8	37.471	58.830	Deciduous broadleaf forest
9	30.522	-18.774	Grassland
10	22.085	7.836	Evergreen broadleaf forest
11	6.544	48.858	Deciduous broadleaf forest
12	-14.821	12.393	Savanna
13	-44.332	-15.083	Savanna
14	-69.863	-48.949	Woody savanna
15	-88.896	39.485	Savanna
16	-105.537	56.866	Wetland

Large variations and fluctuations existed in the original 4-day LAI data in 2003 derived from MODIS product, and some obvious outliers could be easily detected (Figure 3.9). It could be a problem if these LAI data were directly fed into a process-based ecosystem model which used LAI as one of the driving factors, for example the value of LAI could continuously drop and increase by about 6 (Figure 3.9, Site 3). There were three major situations of the noise in MODIS LAI dataset: 1) LAI was underestimated due to the clouds or aerosols in the atmosphere, 2) LAI was overestimated due to the error from the sensors, and 3) LAI highly fluctuated in the typical tropical rainforest regions. The method used in this study to eliminate the outliers removed approximately 15%-48% of the original data (Table 3.3), and was effective and well-performed to all the three situations.

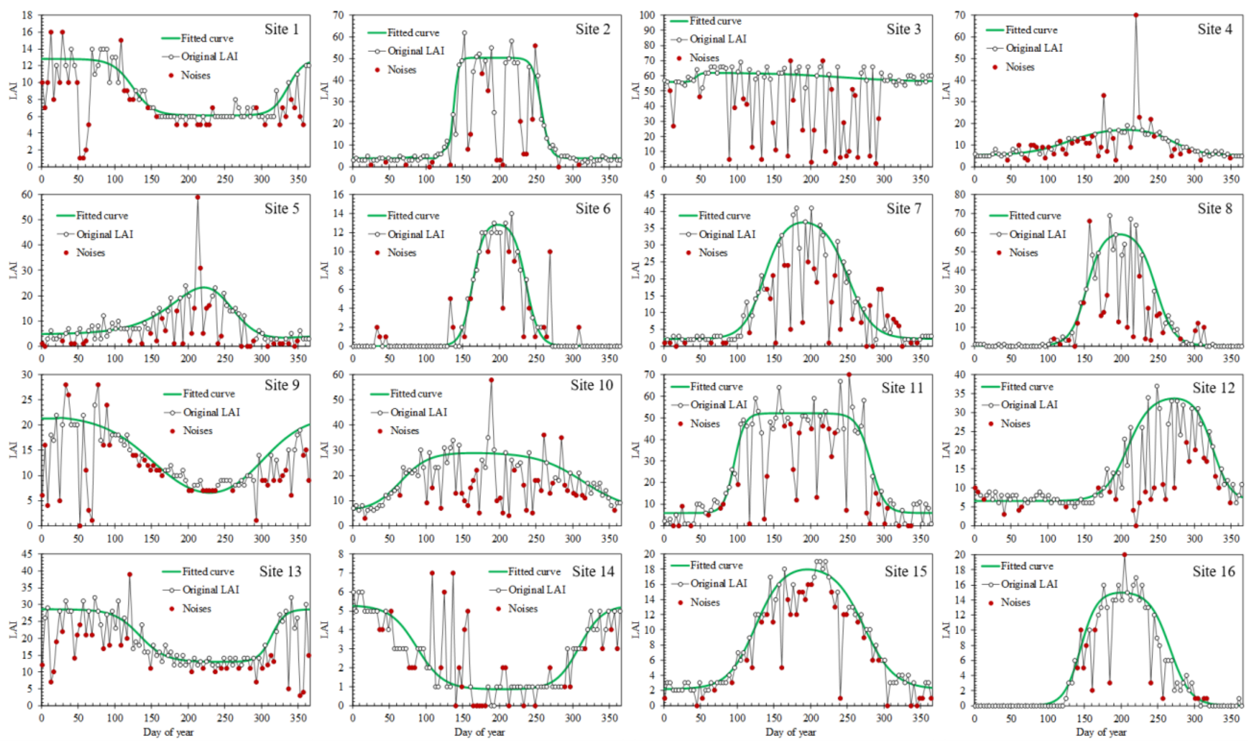


Figure 3.9 The fitted curves (green lines) using double logistic function after removing noise (red dots) of the LAI data derived from MODIS (white dots) for the 16 sample sites with single vegetation cycle in 2003. The values of LAI should be scaled by 0.1.

Table 3.3 Performance of fitting double logistic function to the LAI data after removing noise for the sample sites with single vegetation cycle.

Site ID	Number of data after/before clean	R^2	RMSE
1	74 / 92	0.92	0.084
2	72 / 92	0.97	0.471
3	63 / 92	0.48	0.261
4	55 / 92	0.97	0.088
5	61 / 92	0.86	0.219
6	76 / 92	0.9	0.045
7	51 / 92	0.97	0.253
8	69 / 92	0.92	0.334
9	48 / 92	0.9	0.328
10	56 / 92	0.95	0.280
11	59 / 92	0.81	0.536
12	65 / 92	0.92	0.247
13	61 / 92	0.95	0.216
14	60 / 92	0.77	0.056
15	58 / 92	0.98	0.087
16	78 / 92	0.91	0.112

After eliminating the detected noise and outliers, a double logistic function was fitted to the clean data annually. We found that the fitted curves well characterized the variations and patterns of the annual LAI, reasonably caught the timing of vegetation phenology between growing season and non-growing season, and retained the duration of the peak within the growing season (Figure 3.9). And the fitting was well performed on both southern hemisphere (site 1, 9, 13, and 14) and northern hemisphere (the rest sites). The goodness of fit R^2 of the fitted curves among the 16 sample sites ranged from 0.48 to 0.98 with an average of 0.89 and the RMSE of the fitting ranged from 0.045 to 0.536 with an average of 0.226, indicating a good performance of the double logistic function fitted to the data (Table 3.3). Specifically, the R^2 for the tropical rainforest region was relative lower due to the steady LAI through the year. For example, the lowest R^2 of 0.48 was from site 3, which was dominated by tropical evergreen broadleaf forest and had a stable LAI through the whole year. In addition, many noise existed in site 3 and most of them underestimated the LAI presumably due to cloud contamination. Our approach could accurately detect the noise and derive a steady LAI line after fitting.

3.3.2. Double vegetation cycles

For the grids with two valid vegetation cycles, we selected 8 typical region with different climate and vegetation conditions to evaluate the reconstructing process of LAI in 2003. The sample sites cover 3 major vegetation types, including savanna, grassland, and cropland, and the detailed information of the selected sample sites were shown in Table 3.4.

Table 3.4 Detailed information of the selected sample sites with double vegetation cycles.

Site ID	Longitude	Latitude	Vegetation type
1	115.370	35.030	Cropland
2	115.124	33.030	Cropland
3	83.916	25.362	Cropland
4	44.623	2.914	Cropland
5	-56.354	-21.720	Cropland
6	-57.262	-20.842	Savanna
7	-59.296	-36.879	Grassland
8	-59.468	-36.623	Grassland

Variations and fluctuations also existed in the original 4-day LAI for the grids with double vegetation cycles in 2003 (Figure 3.10). Compared to the noise commonly occurred in the grids with single vegetation cycle, highly fluctuated LAI and overestimated LAI were less remarkable, although underestimation of LAI was still prevalent, especially in growing seasons. The method used in this study to eliminate the noise removed approximately 4%-27% of the original data

(Table 3.5), and was effective and well-performed to both growing season and non-growing season.

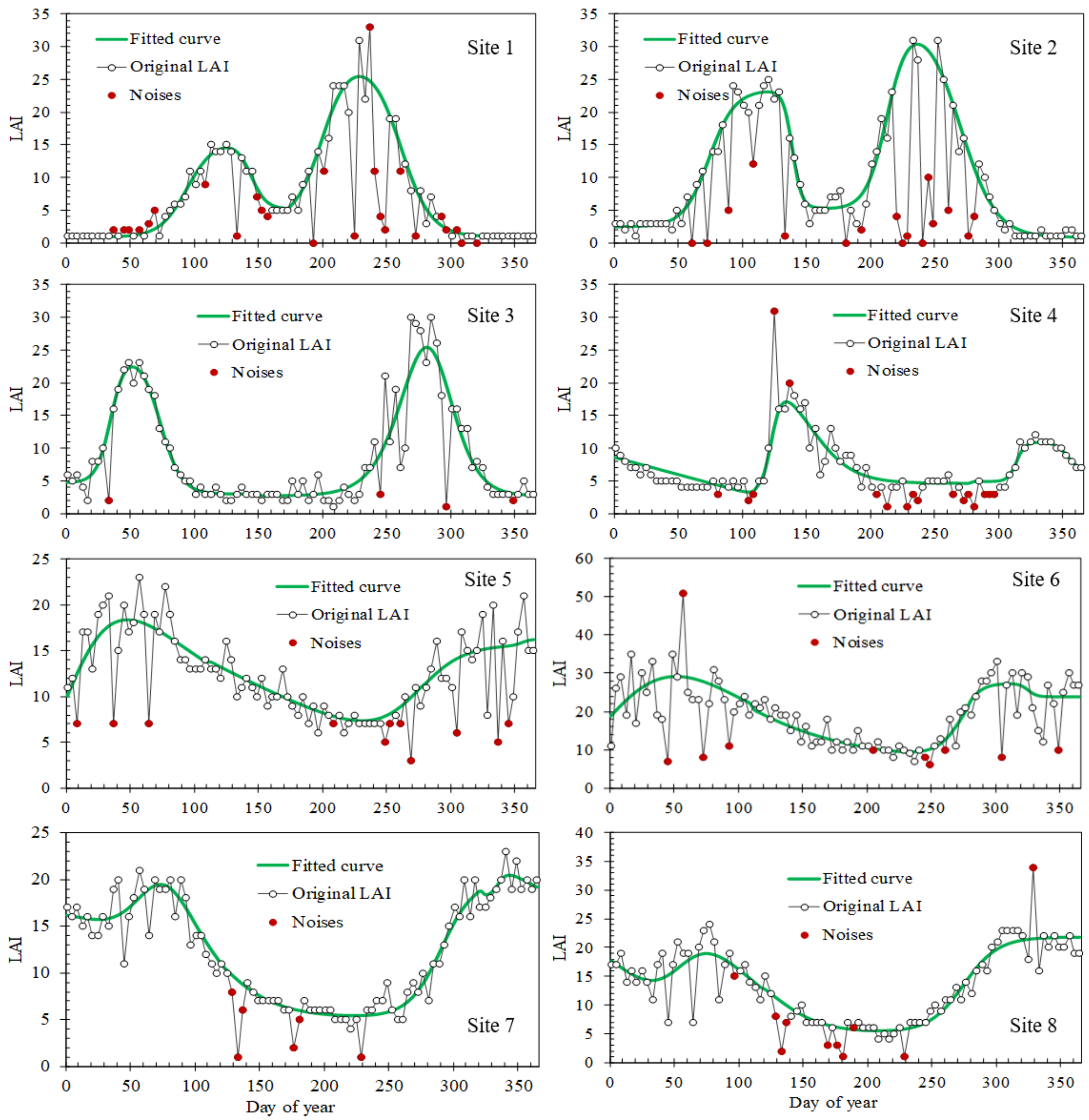


Figure 3.10 The fitted curves (green lines) using double logistic function after removing noise (red dots) of the LAI data derived from MODIS (white dots) for the 8 sample sites with double vegetation cycles in 2003. The values of LAI should be scaled by 0.1.

According to Figure 3.10, the fitted curves of LAI could well characterize the patterns of the double vegetation seasons during a year and reasonably catch the timing of the phenology within and between the two vegetation seasons, and the fitting was well performed on both southern hemisphere and northern hemisphere. The goodness of fit R^2 of the fitted curves among the 8 sample sites ranged from 0.85 to 0.96 with an average of 0.90 and the RMSE of the fitting ranged from 0.141 to 0.484 with an average of 0.232, representing a good performance of the model fitting (Table 3.5). Notably, the fitting performance on the sites with two distinct vegetation seasons (e.g., site 1 and site 3) was relatively better than that on the regions with small LAI variations (e.g., site 5 and site 6) through the year.

Table 3.5 Performance of fitting double logistic function to the LAI data after removing noise for the sample sites with double vegetation cycles.

Site ID	Number of data after/before clean	R^2	RMSE
1	67 / 92	0.92	0.141
2	77 / 92	0.90	0.166
3	88 / 92	0.96	0.261
4	75 / 92	0.87	0.136
5	81 / 92	0.89	0.216
6	82 / 92	0.85	0.484
7	86 / 92	0.87	0.161
8	82 / 92	0.92	0.292

3.3.3. Performance of fitting in global scale

At the global scale, the fitting performance of the LAI data in 2003 was better in boreal, temperate, and part of the tropical regions with R^2 larger than 0.8, and the R^2 was relatively low in most of the tropical region with R^2 less than 0.5 (Figure 3.11). According to the fitting results from the sample sites, although the relatively stable LAI in the tropical region caused the lower

R^2, the fitted curves could reasonably caught the patterns of the LAI (Figure 3.9, site 3). Regarding the vegetation types, the performance was good in most of the vegetation types except evergreen broadleaf forest, such as tropical evergreen broadleaf forest and southeast Asian monsoon rainforest, due to the difficulty of detecting noise and the relatively steady LAI. In general, the method used in this study to reconstruct continuously LAI data, including detecting and removing noise and outliers and fitting double logistic function, performs well globally.

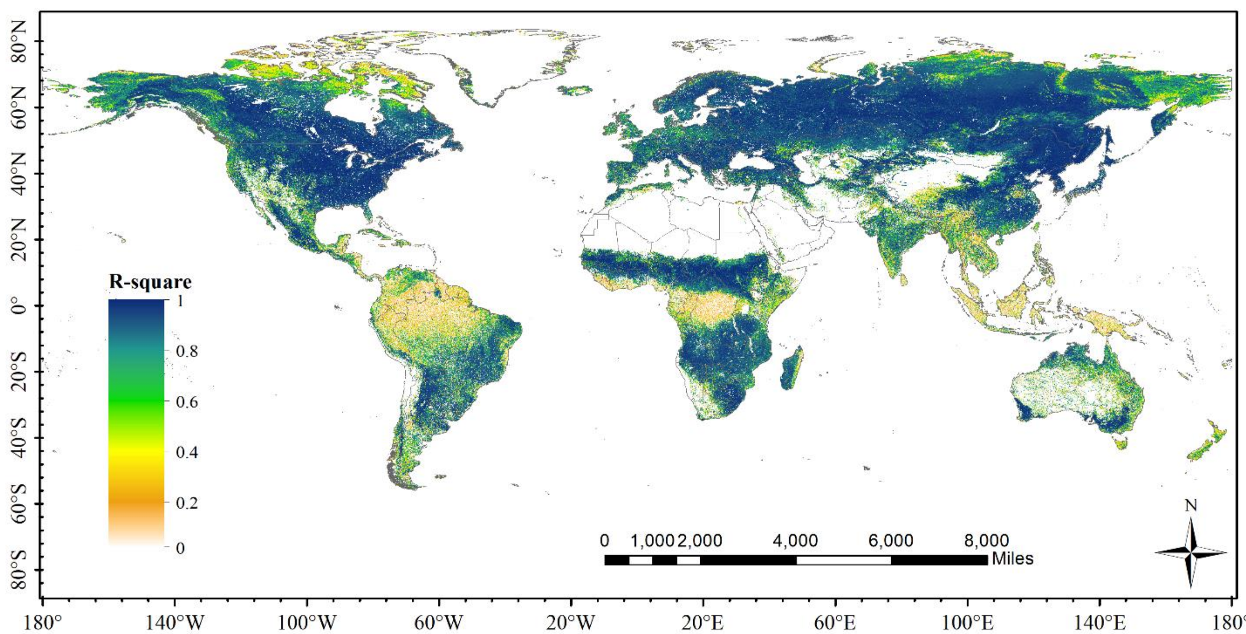


Figure 3.11 Performance of fitting double logistic function of LAI in global scale in 2003.

4. Integrating an improved two-stream canopy radiative transfer model

The performance of simulating canopy radiation absorption and GPP was improved by improving the radiative transfer model with the consideration of radiation partitioning, multiple scattering, and differentiating of sunlit and shaded leaves for photosynthesis. This chapter focused on improving the accuracy of modeling the radiation absorbed by the vegetation canopy and further increasing the performance of terrestrial GPP simulation by integrating a recently developed two stream radiative transfer model that considers multiple scattering in a finite canopy to a two-leaf model. In addition, an empirical radiation partitioning approach was evaluated against 258 site-years from 36 flux tower sites

4.1. Evaluation of radiation partitioning

We first examined the method of partitioning total downward shortwave radiation into direct and diffuse radiation used in our study. Since most flux tower stations lack the records of diffuse radiation, therefore observations of 258 site-years from 36 flux tower sites that provides either diffuse incoming shortwave radiation or diffuse incoming photosynthetic photo flux density information were selected to evaluate the partitioning of radiation. The selected flux towers cover 8 major vegetation types including cropland (CRO), deciduous broadleaf forest (DBF), evergreen broadleaf forest (EBF), evergreen needleleaf forest (ENF), grassland (GRA), mixed forest (MF), open shrublands (OSH), and savanna (SAV), and the information of the flux towers used in this study is shown in Table A.2.

Overall, we found that the modeled diffuse radiation values were in general agreement with the observed diffuse radiation for different vegetation types among the 36 flux tower sites (Figure 4.1). The Pearson's correlation coefficient was larger than 0.75 for all the sites, with an average of 0.87, indicating that the partitioning method used in this study could considerably catch the patterns of direct and diffuse radiation partitioned from total solar radiation. The RMSE varied from 22.89 W/m^2 to 54.41 W/m^2, with the mean of 37.01 W/m^2. The bias ranged from -26.74 W/m^2 to -1.24 W/m^2 with the mean of -12.60 W/m^2. Although the negative biases between our estimated data and observations exhibited a tendency to underestimate the diffuse radiation, the effects were generally small and acceptable. Therefore, the radiation partitioning method used in this study could help to derive the two radiation components: direct and diffuse when only the information of total solar radiation was available.

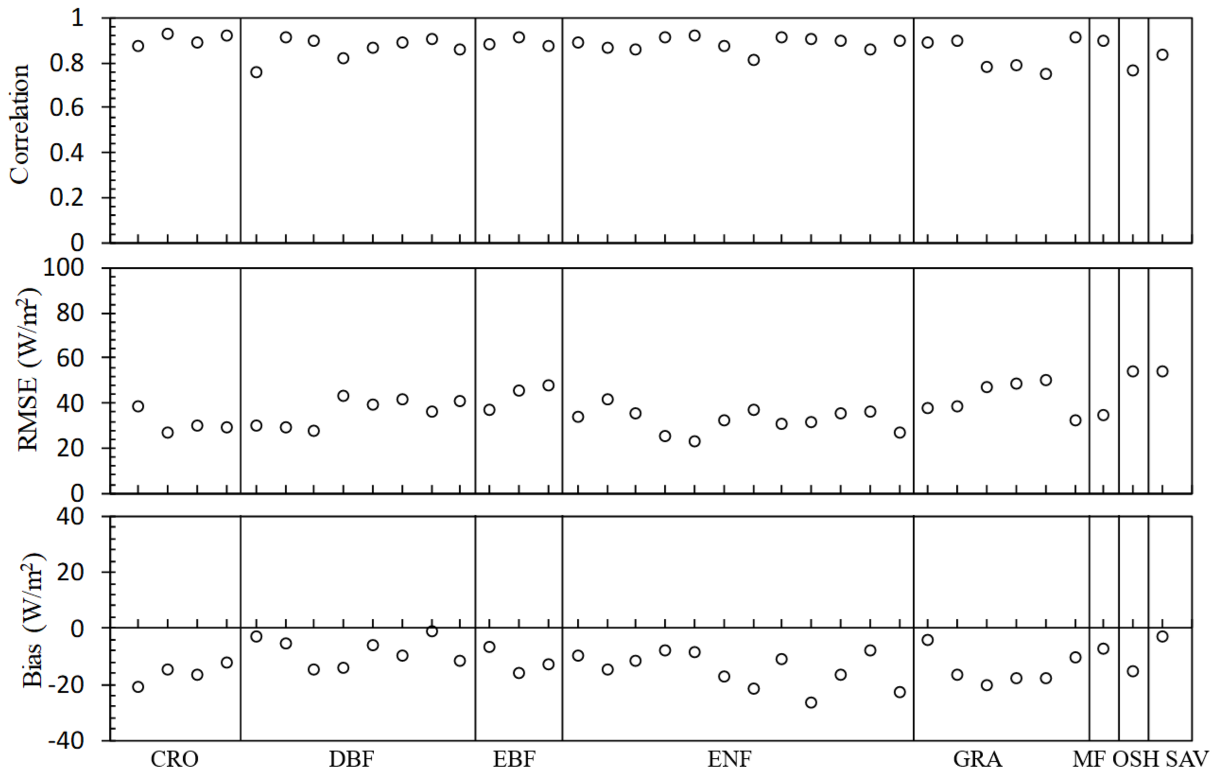


Figure 4.1 Comparison of the observed and modeled diffuse radiation (W/m^2) among the 36 flux tower sites.

4.2. Comparisons of the two stream approach against Beer's law model on radiation absorption by canopy

Beer's law model has been used in many studies to estimate the amount of radiation that vegetation absorbed and various ecosystem models to further quantify vegetation photosynthesis (Wang 2003, Thornton 2010). However, the simple exponential function of Beer's law that does not take scattering effect into account and does not consider direct and diffuse radiation separately might bring large uncertainties when estimating how many solar radiation vegetation canopy can absorb. Figure 4.2 compares the absorbed fraction of solar radiation estimated by Beer's law and the two stream approach used in our study for vegetation canopy with LAI ranging from 0.1 to 8 and diffuse radiation fraction of 0, 0.2, 0.5, 0.8, and 1 under 30° solar zenith angle. It is shown that the absorption fraction simulated by the two stream approach used in our study decreased with the increased diffuse fraction, and the radiation transfer model based on Beer's law had the highest estimates compared to the two stream radiation transfer model (Figure 4.2). Since the Beer's law model did not consider direct and diffuse radiation separately, the curve for absorbed fraction of solar radiation was kept the same under different diffuse radiation fraction levels. The discrepancy in absorption fraction between the two models became larger when the proportion of diffuse radiation increased, and the differences reached up to 73% in an overcast day ($f_b = 1$). This was in agreement with previous studies that Beer's law overestimated the absorbed radiation due to a lack of considering scattering (Wang 2003, Saitoh et al. 2012).

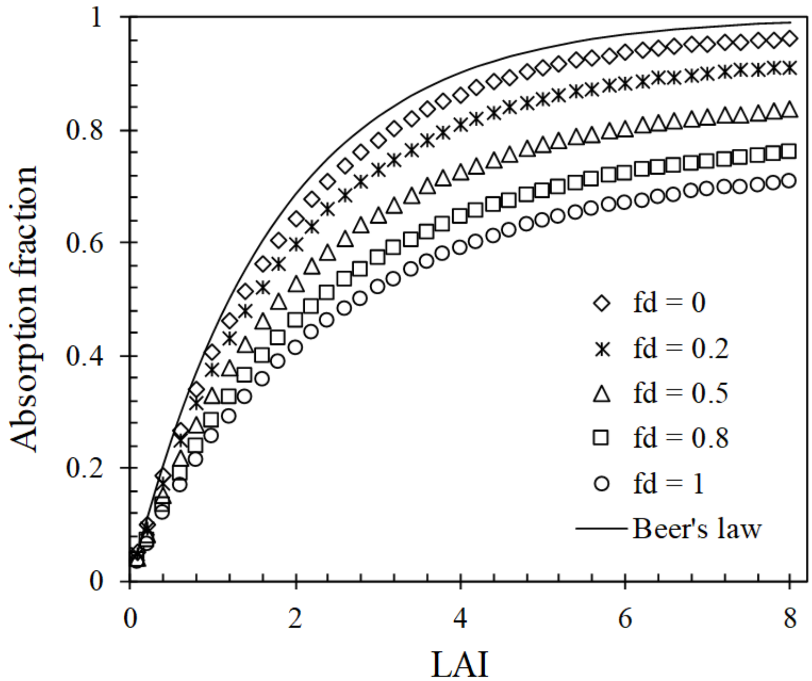


Figure 4.2 Comparison of the absorbed fraction of solar radiation estimated by Beer's law (solid line) and two stream approach (shaped lines) for vegetation canopy with leaf area index ranging from 0.1 to 8 and diffuse radiation fraction (f_b) of 0 (diamond), 0.2 (star), 0.5(triangle), 0.8(square), and 1(circle), respectively. The solar zenith angle (SZA) was set to 30°.

4.3. Improvements of our integrated RTM on estimating GPP

Many previous studies indicated that the sunlit leaves and shaded leaves were different in physiological and biochemical processes due to their different exposure levels to sunlight (de Pury and Farquhar 1997, Zhang et al. 2012, Guan et al. 2022). Sunlit leaves, which receive both direct and diffuse solar radiation, tend to be light saturated for photosynthesis, while the photosynthesis of shaded leaves that absorb the diffuse radiation only is usually limited by the amount of radiation absorbed by leaves. Hence, the big leaf model that does not differentiate sunlit and shaded leaves might induce large biases in estimating canopy radiation absorption and GPP. To accurately estimate the radiation absorbed and the carbon assimilated by the vegetation canopy, we further coupled the two stream radiative transfer model to a two leaf model instead of the big leaf model used in the study of Mahat and Tarboton (2012) and sum the photosynthesis from sunlit and shaded leaves to estimate GPP.

We compared the simulated GPP derived using the Beer's law model (BL), two stream approach with the big-leaf model (TS-BL), and our integrated radiative transfer model (two stream approach with two-leaf model (TS-TL)) with the GPP measurements obtained from six flux tower sites. The selected flux tower sites represented the six most extensive vegetation types in the world, including deciduous broadleaf forest (DBF), evergreen broadleaf forest (EBF), evergreen needleleaf forest (ENF), shrublands (SHB), savanna (SAV), and grassland (GRA), and the information of these flux towers is shown in Table 4.1. The three different radiative transfer models were carried out at a half-hour scale, together with the photosynthesis model by Farquhar et al. (1980), and then we obtained the daily carbon assimilation and compared with the flux tower observations.

Table 4.1 Detailed information of the flux tower sites used to evaluate the radiative transfer model.

Site ID	Vegetation type	Latitude	Longitude	Time Period
US-Wcr	DBF	35.030	115.370	2013-2014
FR-PUE	EBF	33.030	115.124	2007-2008
CA-Obs	ENF	25.362	83.916	2003-2004
CA-NS6	SHB	2.914	44.623	2003-2004
SN-Dhr	SAV	-21.720	-56.354	2012-2013
CN-Dan	GRA	-36.623	-59.468	2004-2005

It is shown in Figure 4.3 that the simulated GPP derived by the three different radiative transfer models (RTM) could generally catch the seasonal and annual patterns during the two years for all the vegetation types though with different variations (Figure 4.3 a1-f1). However, the TS-TL model exhibited most consistent relationships with the flux tower measurements on estimating GPP, followed by the TS-BL model and the BL model, and the corresponding relationships of

TS-TL model were more close to the 1:1 line (Figure 4.3 a2-f2). The RMSE and Bias between the simulated GPP and GPP measurements from flux towers were highest in BL model, and then TS-BL model and TS-TL model for all sites (Table 4.2). The TS-TL model reduced the RMSE and Bias by up to 72% and 81% based on the BL model, and up to 63% and 75% based on the TS-BL model, respectively. The improvements of the TS-TL model were also reflected in the increase in R^2 for all the flux tower sites, and the average R^2 for the TS-TL model was 0.76. In addition, the superior performance of TS-TL model compared to the other two models was more prominent at the sites covered by forests or shrublands and had relatively larger mean GPP, which might be caused by considering multiple scattering and differentiating sunlit and shaded leaves for photosynthesis in dense canopy. In general, our integrated RTM that used TS-TL could distinctly improve GPP estimating compared to that used the TS-BL model and the BL model, especially in the growth seasons.

Table 4.2 The RMSE (gC/m^2/day), Bias (gC/m^2/day), and R^2 between the GPP derived using Beer's law (BL), two-stream approach with big-leaf model (TS-BL), and two-stream approach with two-leaf model (TS-TL) and the GPP measurements from flux tower sites.

	Beer's law			Two stream-BL			Two stream-TL		
Site	RMSE	Bias	R2	RMSE	Bias	R2	RMSE	Bias	R2
US-Wcr	11.76	5.54	0.73	9.02	4.24	0.75	**3.3**	**1.08**	**0.8**
FR-PUE	6.72	4.93	0.38	5.42	3.88	0.39	**2.93**	**1.6**	**0.46**
CA-Obs	1.77	0.41	0.68	1.44	0.08	0.7	**1.26**	**-0.06**	**0.73**
CA-NS6	2.18	0.98	0.86	1.69	0.74	0.87	**0.82**	**0.26**	**0.89**
SN-Dhr	1.95	-0.99	0.73	1.89	-0.98	0.74	**1.6**	**-0.57**	**0.77**
CN-Dan	0.42	0.12	0.86	0.39	**0.09**	0.87	**0.32**	**0.09**	**0.9**

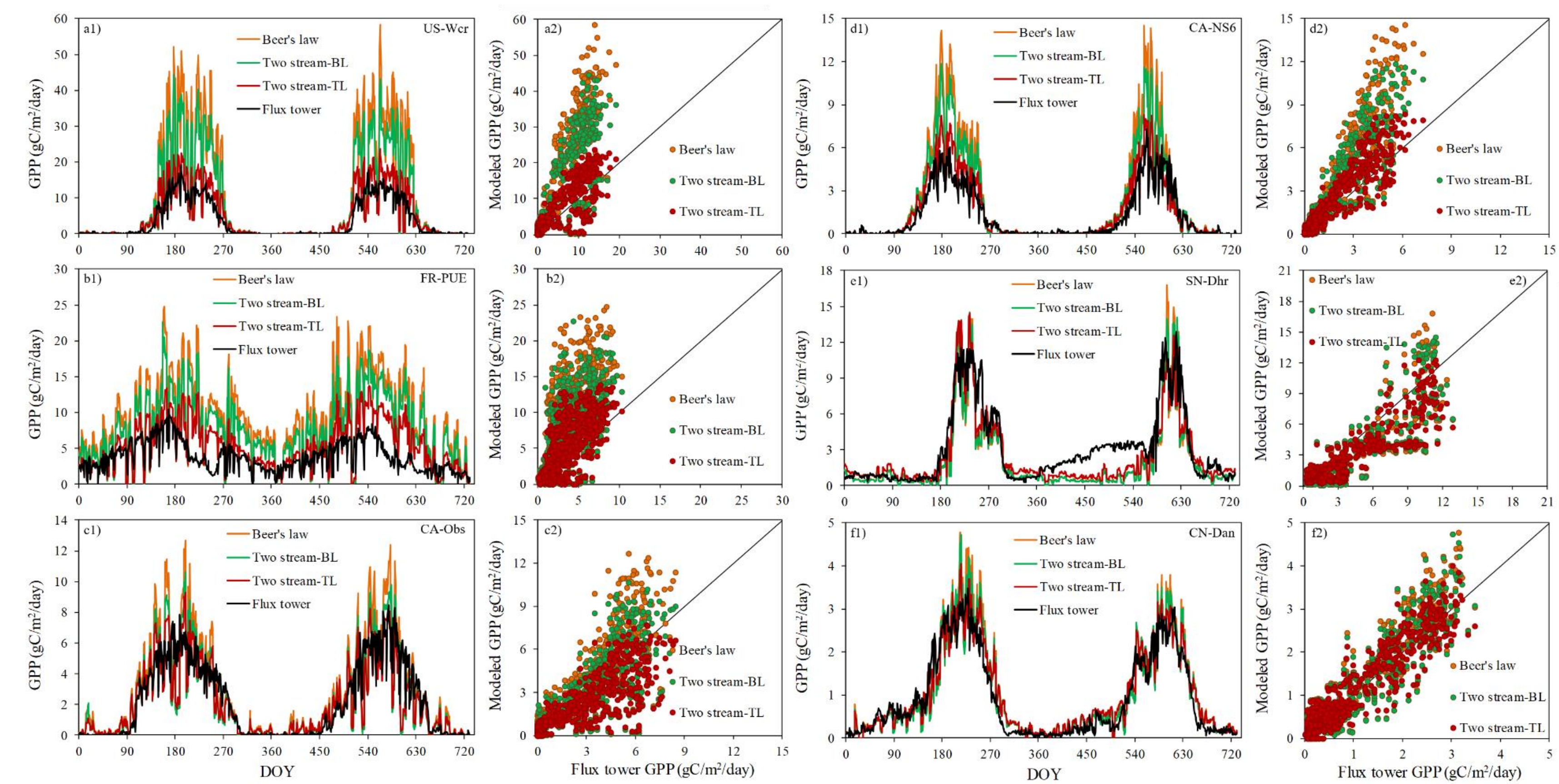


Figure 4.3 Comparisons of the modeled GPP derived by Beer's law (orange), two stream approach with big-leaf model (green), and two stream approach with two-leaf model (red) and the GPP obtained from flux tower observations (black) for sites US-Wcr (a1, a2), FR-PUE (b1, b2), CA-Obs (c1, c2), CA-NS6 (d1, d2), SN-Dhr (e1, e2), and CN-Dan (f1, f2). The vegetation types for US-Wcr, FR-PUE, CA-Obs, CA-NS6, SN-Dhr, and CN-Dan are DBF, EBF, ENF, SHB, SAV, and GRA, respectively. The black lines in a2-f2 are the 1:1 lines.

We further investigated the performance of our integrated RTM that used the TS-TL model in improving GPP estimations for different LAI levels. The daily GPP differences between the simulated results for the three different RTMs and the flux tower observations among the six flux towers during the two years were classified into 7 LAI groups based on LAI values. The boxplots in Figure 4.4 showed that the improvements in GPP estimations by using our integrated TS-TL model compared to using the other two models were more significant in larger LAI groups that represented dense vegetation canopy coverage, while the three RTMs performed similarly well in simulating GPP at sparse vegetation coverage or non-growing seasons where the values of LAI were small. The GPP differences displayed by the TS-BL model were smaller than that simulated by BL model with LAI ranging from 2.1 to 6, but they were almost the same when LAI became larger. Multiple scattering could increase the radiation transmitted out the canopy, and thus reduce the absorbed radiation by the vegetation canopy (Nijssen and Lettenmaier 1999, Zhao and Qualls 2005). However, when the canopy became too thick, the radiation that penetrated below the canopy approached zero even considering multiple scattering, which might explain the ineffectiveness of the TS-BL model at excessively thick vegetation canopy. Nevertheless, differentiating sunlit and shaded leaves lessened the biases when estimating canopy photosynthesis due to the non-linear effects of absorbed radiation on the partitioning of available energy and photosynthesis, which was confirmed by previous studies (Wang and Leuning 1998, Dai et al. 2004). The considerable improvements at larger LAI values by using the TS-TL model indicated that with the considering of multiple scattering or differentiating sunlit and shaded leaves largely reduced the biases in modeling the light profiles and estimating total photosynthesis by vegetation. Therefore, our integrated RTM (TS-TL model) exhibited relatively robust and consistent performance in accurately simulating carbon assimilation for different LAI levels.

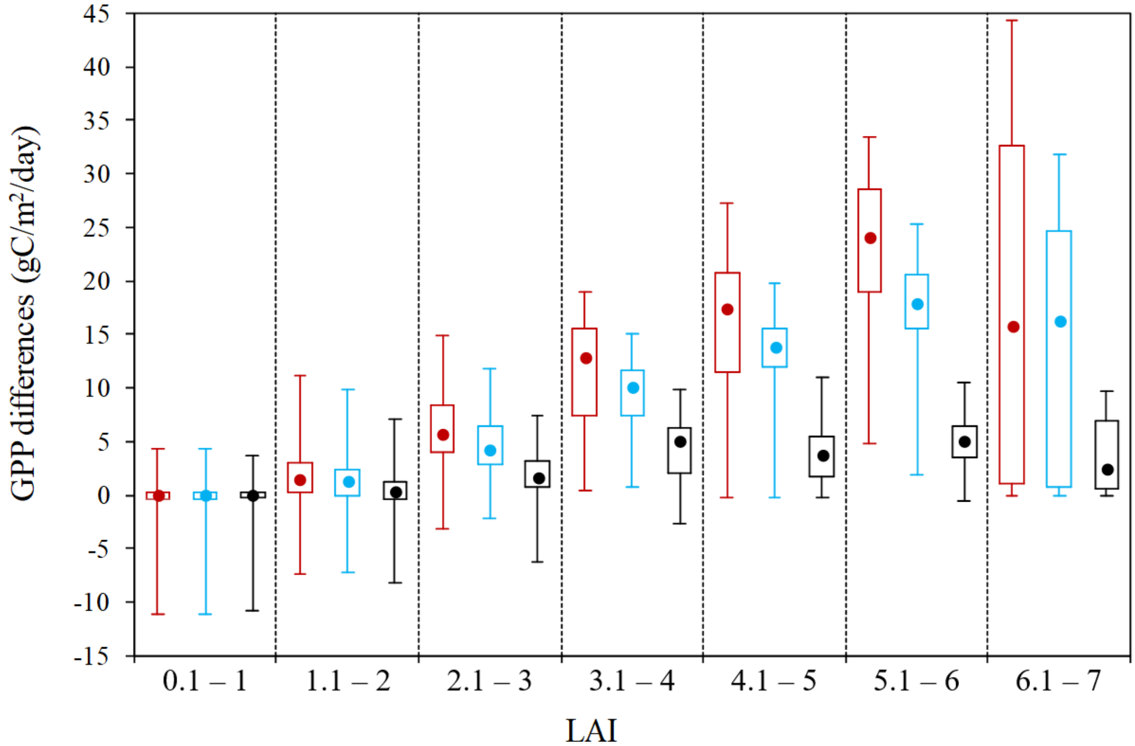


Figure 4.4 Boxplots of GPP differences between the simulated GPP derived by Beer's law (red), two stream approach with big-leaf model (blue), and two stream approach with two-leaf model (black) and the GPP measurements from flux tower sites.

5. Mapping global terrestrial gross primary productivity from local sites to global values using an improved process-based ecosystem model on Google Earth Engine Platform during 2001-2020

An improved process-based ecosystem model driven by satellite-based LAI data and reanalysis climate data was developed to achieve the terrestrial GPP estimation at global scale using high performance computing during the past two decades. In this chapter, a comprehensive process-based model that coupled the improved two stream radiation transfer model was developed on the Google Earth Engine platform, and the simulated GPP was evaluated against 167 flux tower sites. The spatial and temporal patterns and trends in global terrestrial GPP during 2001-2020 were examined for different vegetation types, and the comparisons of global GPP estimates from recent studies and products were discussed. In addition, the sensitivities of our model to environmental and biological drivers were also investigated in this chapter.

5.1. Comparison against flux tower sites

Our modeled gross primary productivity results were evaluated against eddy covariance flux towers from the FLUXNET2015 Dataset. After a rigorous data quality check, we removed the flux towers with inconsistent time series and too many missing or poor quality values, and 167 flux tower sites across the globe were finally selected in our study. The selected flux towers cover 12 major vegetation types, including evergreen needleleaf forests (ENF), evergreen broadleaf forests (EBF), deciduous needleleaf forests (DNF), deciduous broadleaf forests (DBF), mixed forests (MF), closed shrublands (CSH), open shrublands (OSH), woody savannas (WSA), savannas (SAV), grasslands (GRA), wetlands (WET), and cropland (CRO), and the information of the flux tower sites used in this study was shown in Table A.3. For each site, the half-hourly air temperature, incoming shortwave radiation, atmospheric pressure, and VPD were used to drive our integrated process-based model, and the half-hourly GPP data was used to evaluate our model output. The validation was carried out at both half-hour and annual scales: 1) firstly we directly compared the half-hourly GPP estimates for all sites and then 2) we calculated the annual average GPP from both flux tower observations and model outputs for evaluating at annual scale.

According to Figure 5.1, our integrated process-based model performed well in reproducing temporal variations in GPP at most flux tower sites, and our modeled GPP estimates were highly correlated to the flux tower observations at a half-hourly scale. The Pearson's correlation coefficient between our modeled GPP and flux tower observations varied from 0.39 at the GH-Ank site to 0.94 at the RU-Ha1 site, and the mean correlation coefficient was 0.77 over all flux tower sites. The relative lower correlation coefficients ($r < 0.5$) were found at tropical forests (e.g., GH-Ank with $r = 0.39$) or sparse vegetation areas (e.g., ES-LJu site with $r = 0.48$, CN-Du3 site with $r = 0.49$). The averaged RMSE across all the sites was 3.54 $gC/m^2/day$, ranging from 0.75 $gC/m^2/day$ to 9.51 $gC/m^2/day$. And the bias varied from -2.05 $gC/m^2/day$ to 4.62 $gC/m^2/day$ with the mean of 0.79 $gC/m^2/day$. In general, the high correlation and low RMSE and bias indicated an overall high accuracy of our integrated process-based model in simulating GPP at half-hourly scale.

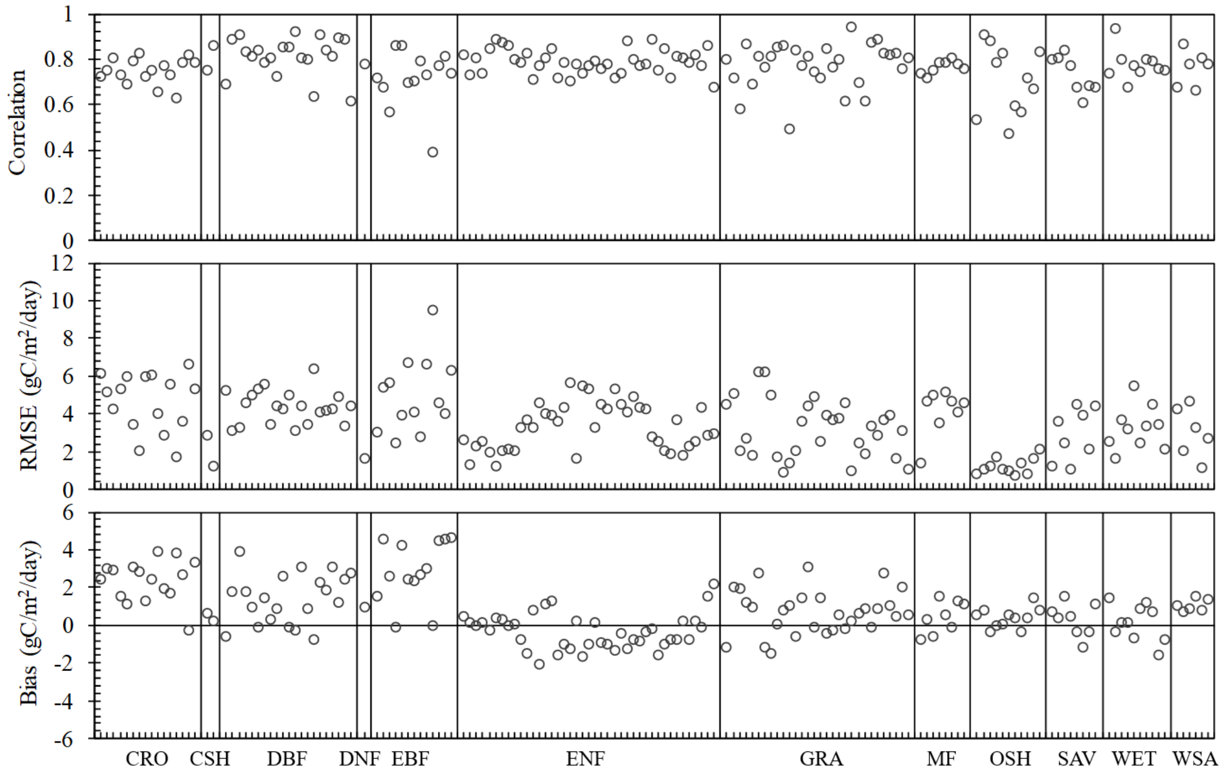


Figure 5.1 Comparison of our model-simulated GPP against 167 flux tower-based GPP.

For individual vegetation types, our model-simulated GPP was highly correlated to the flux tower-based GPP for the 12 vegetation types, with all the average correlation coefficients larger than 0.71 (Figure 5.2). The correlation coefficient was highest in DBF (r = 0.82), while it was lowest in OSH (r = 0.71). The values of RMSE and bias were both highest in EBF with 5.02 gC/m^2/day and 2.85 gC/m^2/day respectively, where the forests were mainly distributed in tropical regions. The complex canopy structures might bring uncertainties to GPP estimating (Zhang et al. 2019), therefore considering the canopy structures such as distributions of leaf angle and clumping factors in the process-based model is suggested to improve the estimation of tropical forest carbon cycling in the future. Although positive biases were found in all the vegetation types except ENF, indicating that our integrated process-based model has slightly overestimated

GPP, the small values of bias were acceptable to us. In general, our model could produce reliable estimations of GPP in most vegetation types.

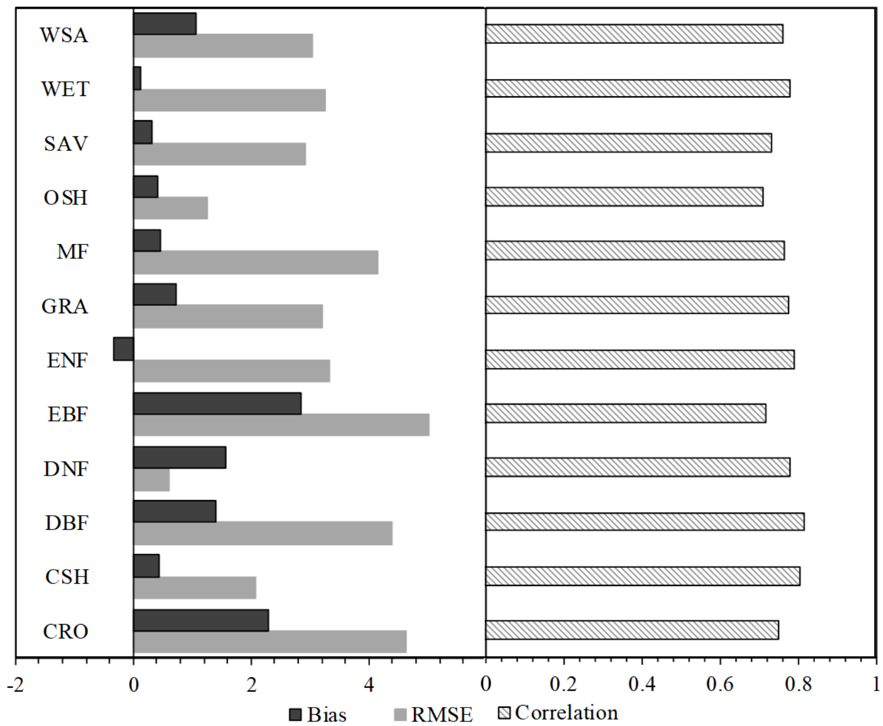


Figure 5.2 Correlation coefficient, RMSE (gC/m^2/day), and bias (gC/m^2/day) between our model-simulated and flux tower-based GPP for different vegetation types.

We also evaluated our model-simulated GPP against the observations from 167 flux tower sites at an annual scale (Figure 5.3). The annual mean GPP derived from our integrated process-based model showed a strong significant linear relation ($p < 0.01$) with the flux tower data, and it explained 72% of the spatial variations in GPP across all the validation sites. The small values of relative RMSE and bias between our model-simulated and flux tower-based annual mean GPP, with 28% and 1% respectively, indicating that our integrated process-based model performed well in estimating GPP at annual scale with a slightly overestimation.

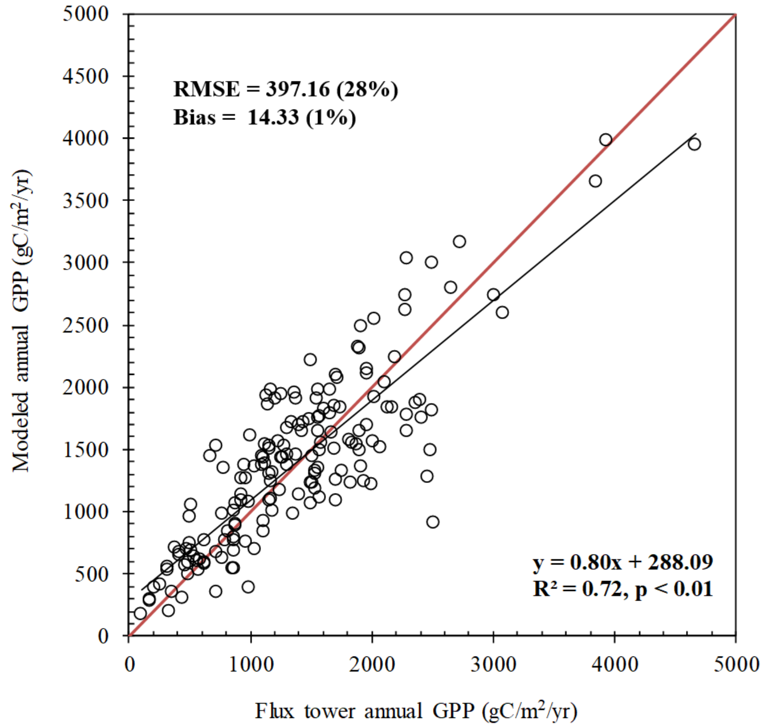
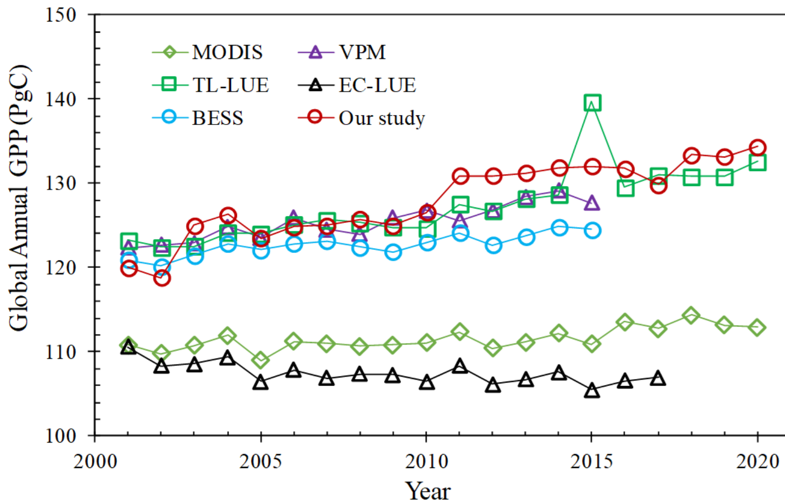


Figure 5.3 Comparison of our model-simulated annual mean GPP and flux tower-based annual GPP. The black line is the regression line, and the red line is the 1:1 line. Unit: $gC/m^2/day$.

5.2. Temporal and spatial variations in global terrestrial GPP

Global terrestrial GPP at 0.25° spatial resolution and 3-hourly temporal resolution was simulated using our integrated process-based model from 2001 to 2020 driven by reconstructed LAI data from MODIS MCD15A3H and climate data from GLDAS 2.1. The annual global terrestrial GPP ranged from 118 PgC to 134 PgC, with an average of 128 PgC, during the study period from 2001 to 2020 (Figure 5.4). And it showed a significant increasing trend ($p < 0.01$) during the two decades, with an average rate of increase of 0.71 PgC/yr globally. We found that the mean of the annual GPP after 2010 is 132 PgC/yr, which is 6% higher than the mean of the annual GPP before 2010 (124 PgC/yr). Interestingly, an elevation after 2010 was also found in solar radiation

(Figure 3.7), which acted as major energy input for vegetation photosynthesis, and the higher solar radiation after 2010 might be responsible for the higher annual GPP.



Model	Linear regression	R^2	p-value
MODIS	$y = 0.17x - 224.99$	0.54	<0.01
VPM	$y = 0.43x - 735.20$	0.84	<0.01
TL-LUE	$y = 0.58x - 1034.09$	0.66	<0.01
EC-LUE	$y = -0.17x + 455.20$	0.48	<0.01
BESS	$y = 0.25x - 371.50$	0.72	<0.01
Our study	$y = 0.71x - 1299.26$	0.86	<0.01

Figure 5.4 Time series of global annual GPP from our study and from a set of previous estimates. The right table shows the corresponding linear regressions.

Annual GPP also varied in different vegetation types from 2001 to 2020 (Table 5.1), and the annual average GPP from largest to lowest were 2.61 KgC/m^2/yr in EBF, 1.29 KgC/m^2/yr in DBF, 1.07 KgC/m^2/yr in WSA, 1.02 KgC/m^2/yr in MF, 0.99 KgC/m^2/yr in CRO, 0.91 KgC/m^2/yr in SAV, 0.77 KgC/m^2/yr in DNF, 0.77 KgC/m^2/yr in ENF, 0.67 KgC/m^2/yr in WET, 0.44 KgC/m^2/yr in GRA, 0.38 KgC/m^2/yr in CSH, and 0.20 KgC/m^2/yr in OSH. In addition, the significant increasing trend ($p < 0.01$) in annual GPP was found in all the vegetation types except in EBF ($p = 0.05$), and the percentages of increase were 24.53% in ENF, 2.36% in DNF, 25.62% in DBF, 12.29% in MF, 19.14% in CSH, 12.47% in OSH, 42.19% in WSA, 13.60% in SAV, 14.04% in GRA, 17.97%, 14.84% in WET, and 18.18% in CRO.

Table 5.1 Annual GPP ($KgC/m^2/yr$) among different vegetation types during 2001-2020. (ENF: evergreen needleleaf forests, EBF: evergreen broadleaf forests, DNF: deciduous needleleaf forests, DBF: deciduous broadleaf forests, MF: mixed forests, CSH: closed shrublands, OSH: open shrublands, WSA: woody savannas, SAV: savannas, GRA: grasslands, WET: wetlands, CRO: croplands).

Year	ENF	EBF	DNF	DBF	MF	CSH	OSH	WSA	SAV	GRA	WET	CRO	Global
2001	0.66	2.55	0.71	1.23	0.94	0.35	0.16	0.99	0.84	0.39	0.61	0.91	0.89
2002	0.67	2.51	0.74	1.21	0.93	0.34	0.17	0.99	0.84	0.40	0.60	0.89	0.88
2003	0.74	2.62	0.70	1.23	0.99	0.36	0.18	1.04	0.88	0.42	0.68	0.93	0.93
2004	0.73	2.66	0.68	1.26	0.97	0.38	0.18	1.03	0.88	0.43	0.64	0.97	0.94
2005	0.75	2.56	0.73	1.23	0.99	0.35	0.18	1.03	0.87	0.42	0.67	0.94	0.92
2006	0.75	2.59	0.71	1.25	0.96	0.38	0.18	1.04	0.88	0.42	0.67	0.93	0.93
2007	0.72	2.59	0.72	1.25	0.98	0.35	0.19	1.02	0.89	0.43	0.65	0.95	0.93
2008	0.70	2.65	0.73	1.26	0.95	0.38	0.18	1.02	0.89	0.42	0.64	0.96	0.93
2009	0.72	2.63	0.73	1.25	0.96	0.39	0.18	1.02	0.88	0.42	0.63	0.93	0.93
2010	0.74	2.61	0.74	1.27	1.00	0.40	0.20	1.04	0.89	0.44	0.65	0.96	0.94
2011	0.79	2.64	0.85	1.32	1.07	0.42	0.21	1.09	0.93	0.45	0.69	1.03	0.97
2012	0.80	2.65	0.83	1.35	1.07	0.40	0.22	1.10	0.93	0.45	0.71	1.01	0.97
2013	0.83	2.65	0.81	1.31	1.06	0.40	0.21	1.10	0.93	0.45	0.69	1.02	0.97
2014	0.82	2.64	0.83	1.33	1.07	0.40	0.21	1.10	0.94	0.45	0.68	1.04	0.98
2015	0.84	2.65	0.83	1.35	1.09	0.38	0.21	1.11	0.94	0.45	0.70	1.03	0.98
2016	0.87	2.57	0.82	1.34	1.13	0.40	0.23	1.12	0.95	0.47	0.75	1.06	0.98
2017	0.81	2.57	0.81	1.34	1.04	0.40	0.22	1.09	0.93	0.46	0.69	1.03	0.96
2018	0.85	2.65	0.85	1.36	1.10	0.38	0.22	1.13	0.96	0.46	0.70	1.04	0.99
2019	0.81	2.62	0.85	1.34	1.07	0.37	0.23	1.12	0.96	0.47	0.70	1.05	0.99
2020	0.82	2.65	0.86	1.34	1.11	0.40	0.23	1.12	0.96	0.48	0.73	1.07	1.00
Mean	0.77	2.61	0.77	1.29	1.02	0.38	0.20	1.07	0.91	0.44	0.67	0.99	0.95

The spatial patterns of annual mean GPP simulated by our integrated process-based model during 2001 to 2020 was shown in Figure 5.5a. The highest annual GPP occurred mainly in the tropical regions, especially in the evergreen broadleaf forests in Amazon and Southeast Asia, while the lowest GPP was mostly distributed in cold or arid areas. During the period of 2001-2020, approximately 83.8% of global terrestrial production showed an increasing trend in annual

GPP, where more than half of these regions were statistically significant at a 95% confidence level, and these significant increasing trends were mainly distributed in temperate regions and high elevation regions (Figure 5.5b). Only 3.2% of the global terrestrial production showed significantly decreasing trends in annual GPP, which were scattered and located in the tropical rainforests such as Congo Basin and the Amazon rainforests. These spatial patterns in annual mean GPP and its trends with extended updates of GPP estimates to 2020 were consistent with previous studies (Zhang et al. 2017, Zheng et al. 2020).

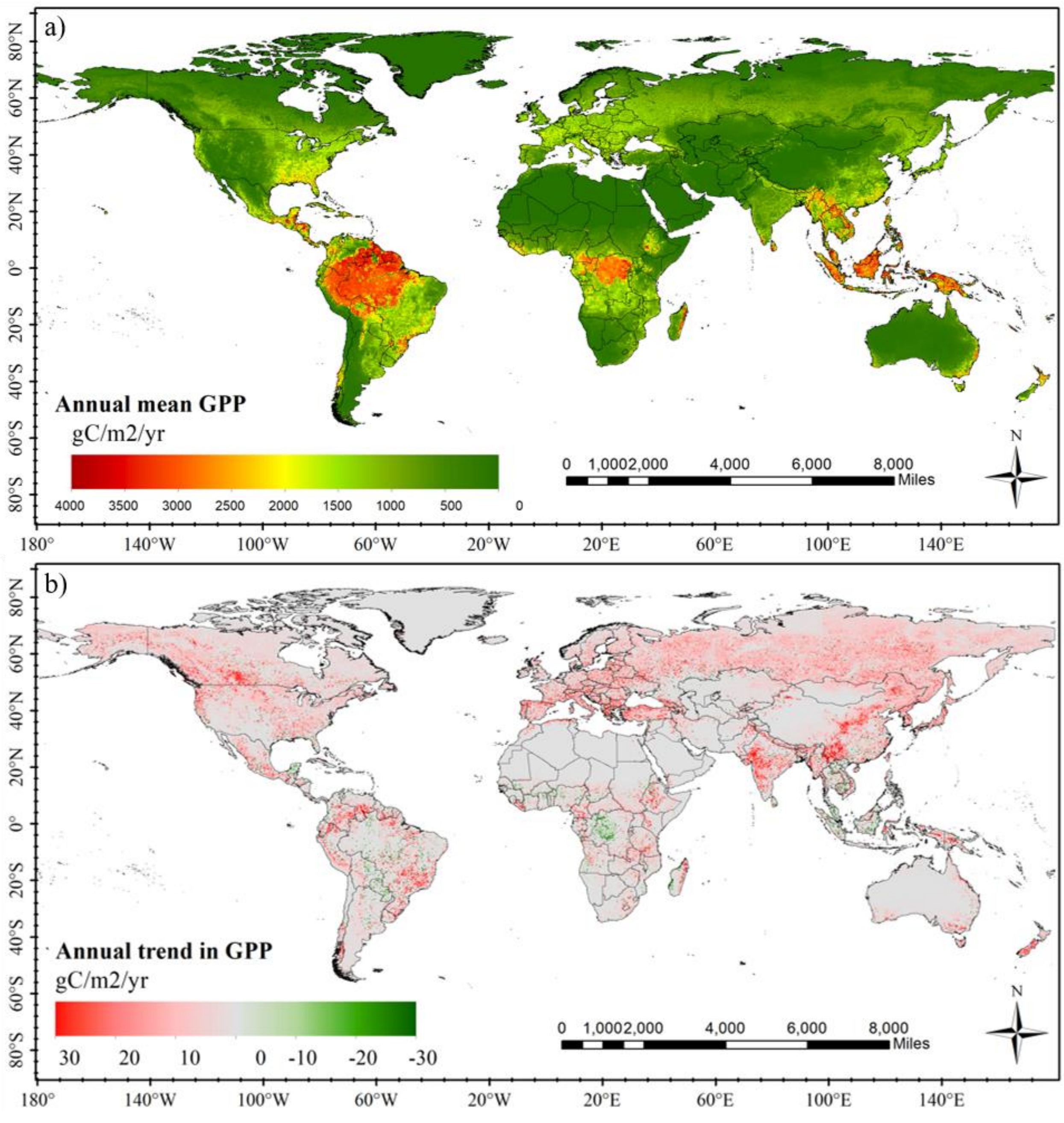


Figure 5.5 Spatial distributions of global a) annual mean GPP and b) significant trend ($p < 0.05$) of GPP during the period of 2001-2020. (Unit: $gC/m^2/yr$)

5.3. Global estimates of terrestrial GPP

Despite the existence of many different methods and models to quantify global terrestrial GPP, it is still subject to uncertainty caused by different sources of input data, settings of parameters, and model structures. The global estimates of terrestrial GPP in the recent two decades are reported varying between 105 PgC/yr and 140 PgC/yr (Zhao and Running 2010, Jiang and Ryu 2016, Zhang et al. 2017, Zheng et al. 2020, Bi et al. 2022) using different simulation methods (Figure 5.4). The GPP simulations of MODIS (MOD17A2H), VPM, TL-LUE, and EC-LUE model are all based on light use efficiency model, and the averaged estimates in GPP are 112 PgC/yr (2001-2020), 125PgC/yr (2001-2015), 127 PgC/yr (2001-2020), and 107 PgC (2001-2017), respectively. Averaged global annual GPP estimated by BESS, a process-based model, is 123 PgC/yr from 2001 to 2016. We quantified the averaged global annual GPP from 2001 to 2020 as 128 PgC/yr, which is comparable to the previous estimates. In addition, our simulated GPP showed a significantly increasing trend, and the increase is consistent with most other GPP estimations excepting EC-LUE who showed a significant declining trend. However, the increase rate of 0.71 PgC/yr from our study is a little larger than others due to the higher GPP after 2010. We also compared our model simulation with MODIS (MOD17A2H) in annual terrestrial GPP averaged over the period of 2001-2020 spatially (Figure 5.6). In general, our model showed higher annual GPP estimates in forests, especially in the tropical rainforest regions that are mostly dominated by evergreen broadleaf forest. This might be caused by the different model structure used to simulate GPP. The MODIS GPP was estimated based on light use efficiency model, while our integrated process-based model considered a complex two leaf canopy radiative transfer model which could more accurately quantify how much radiation plants absorb for photosynthesis use especially for the vegetation areas with large LAI. Meanwhile, our simulated GPP was lower in savanna and open shrublands, where the vegetation coverage is relatively low. However, our results for the average annual GPP in forests was more consistent with the BESS result (Jiang and Ryu 2016), which might confirms the advantages of using process-based model with modelling the underlying mechanisms within canopy in mapping GPP.

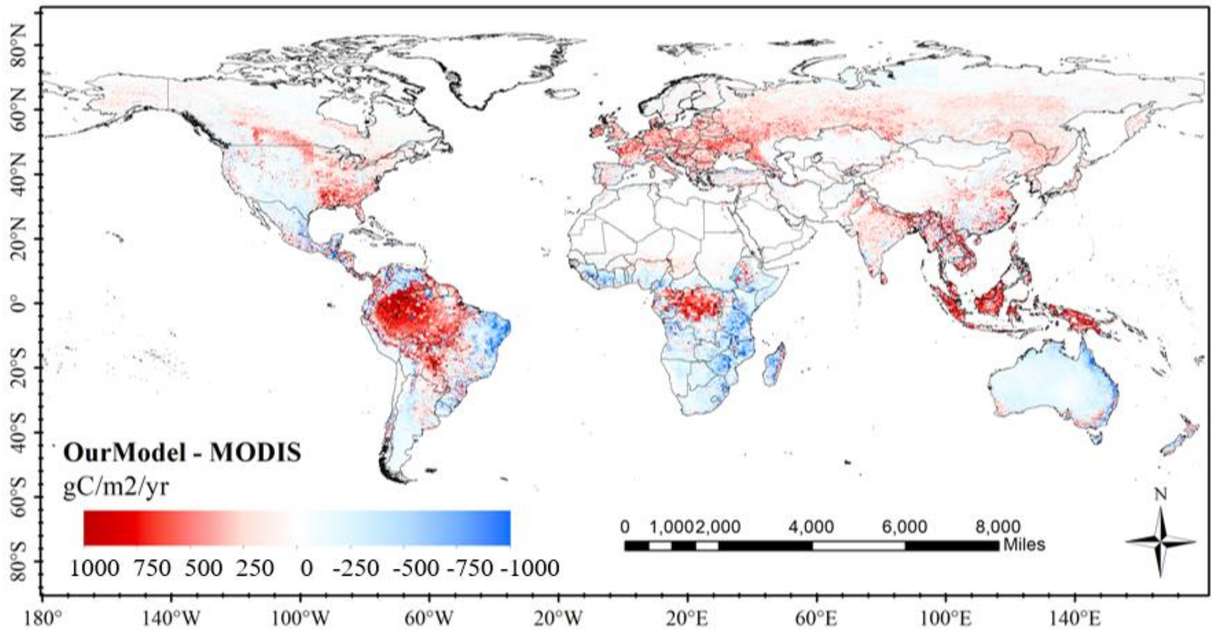


Figure 5.6 Difference map of annual mean GPP between our study and MOD17 for the period of 2001-2020. (Unit: gC/m^2/yr)

5.4. Sensitivity analysis on estimated GPP

We performed a simple sensitivity analysis for the global GPP outputs from our model in 2013, and six key environmental and biophysiological variables, including air temperature, solar radiation, atmospheric CO_2 concentration, vapor pressure deficit (VPD), leaf area index (LAI), and V_{cmax25}, were selected to examine the sensitivities of our model to these drivers. According to the result, GPP derived from our model were more sensitive to biophysiological variables compared to the environmental variables (Figure 5.7). GPP was most sensitive to V_{cmax25}, and then LAI, atmospheric CO_2 concentration, solar radiation, air temperature, and VPD. A 30%

change in V_{cmax25}, LAI, CO_2 concentration, solar radiation, air temperature, and VPD led to approximately 21.18%, 21.57%, 18.98%, 14.32%, 10.12%, and 2.69% change in GPP, respectively.

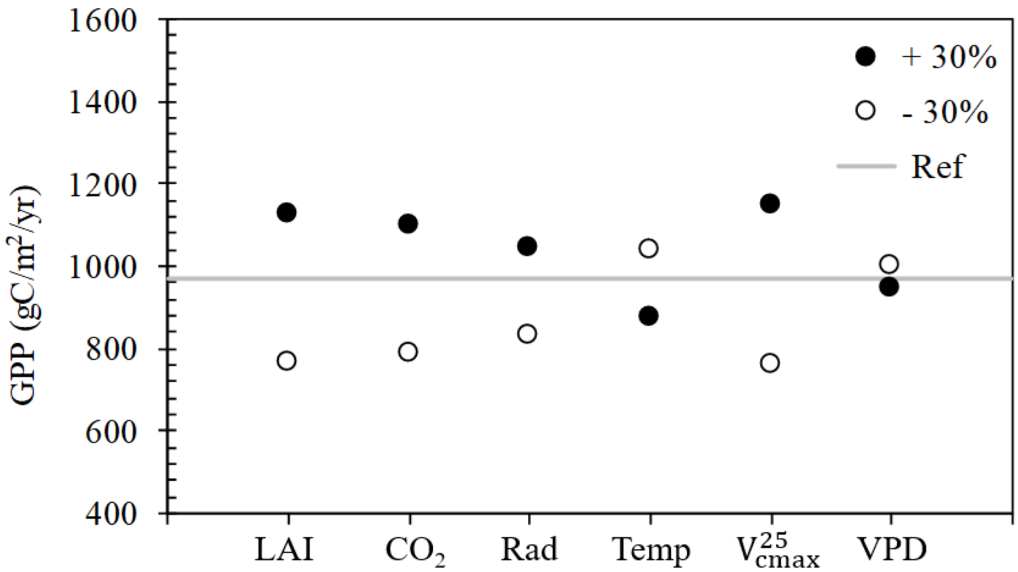


Figure 5.7 Sensitivity analysis of our model. Each of the six variables (LAI: leaf area index, CO_2: atmospheric CO_2 concentration, Rad: downward shortwave radiation, Temp: air temperature, V_{cmax}^{25}: maximum rate of carboxylation at 25 degree C, and VPD: vapor pressure deficit) was changed ±30% with keeping everything else the same, and the GPP (gC/m^2/yr) results were compared.

For process-based ecosystem models, biophysiological variables, as essential model parameter inputs, directly influence the estimation of carbon fluxes. Variation in biophysiological factors could lead to large uncertainties in quantifying terrestrial carbon cycle (Zaehle et al. 2005, Kala et al. 2014, Rogers 2014, Walker et al. 2017, Liu et al., 2018, Xie et al. 2019). Our results showed that our model derived GPP was most sensitive to biophysiological parameters V_{cmax25} and LAI, which was in agreement with the findings reported by Ryu et al. (2011) that BESS-derived GPP was most sensitive to LAI and V_{cmax25} with about 25% and 15% changes in GPP when LAI and V_{cmax25} changed 30% in 2003. Also, Bonan et al. 2011 found that model uncertainty over V_{cmax} has been shown to account for about 30 PgC/yr variation in estimation of GPP. However, accurate biophysiological parameters are difficult to obtain, especially in large scale, due to the specificity of different vegetation types and even different plant species, which might cause a large discrepancy on estimating global carbon cycle (Roger et al. 2017). Hence, obtaining accurate biophysiological parameters at a large scale is challenging but necessary and crucial to improve the accuracy of modeling the ecosystem carbon cycle in the future.

6. Summary

6.1. Summary

This dissertation provides insights with respect to achieving an improved global terrestrial GPP simulation based on an integrated process-based model on the Google Earth Engine cloud platform during 2001-2020 through increasing the performance of the radiation transfer model and reconstructing remote sensing LAI products, which is critical to better understanding of global carbon cycle in terrestrial ecosystems. The main findings of this study are presented below in this chapter.

In chapter 3, five primary climate variables, including air temperature, precipitation, downward shortwave radiation, air pressure, and VPD (derived from specific humidity), obtained from GLDAS 2.1 were evaluated against the observations from 167 worldwide flux tower sites that cover 13 major vegetation types at both daily and annual scales. All of the GLDAS-derived climate variables showed strong correlations compared to the flux tower observations at daily scale with the average Pearson's correlation coefficients of 0.97, 0.91, 0.91, and 0.86 for temperature, solar radiation, air pressure, and VPD, respectively, except precipitation whose average correlation coefficient was 0.44. The GLDAS 2.1 precipitation data exhibited an overall overestimation with 85% sites showing positive bias, while other four variables showed small negative average biases. Annually, though significant linear relationships were found for all of the five climate variables, the large relative RMSE and relative bias in precipitation revealed that the GLDAS 2.1 tended to overestimate the precipitation with an average bias of 222.22 mm. Overall, the GLDAS 2.1 performed reliable estimations in air temperature, downward shortwave radiation, air pressure, and VPD globally, besides a large overestimation in precipitation data. In addition, the temporal and spatial patterns of these variables were examined during the period of

2001-2020. The global air temperature, solar radiation, VPD, and precipitation showed a significantly increasing trend with rates of 0.7°C/decade, 3.1W/m^2/decade, 0.15KPa/decade, and 49.6mm/decade, respectively, while air pressure did not show any significant changes from 2001 to 2020. The five climate variables showed different spatial variations globally and their changes over the past decades were not homogenous in space. During the recent two decades, the air temperature and VPD showed significant warming trend in most tropical and subtropical regions, and the most pronounced warming and drying trend appeared at the high elevation regions. Solar radiation significantly increased in the regions near 70°N in latitude, most Africa, the northern South America, the southern India, and the western Australia, while eastern China, some regions in Europe and the middle north of America showed a significantly dimming trend during the twenty years. It was interesting to find that the regions with significantly decreasing solar radiation were also where aerosols most affected areas in recent decades, indicating the important role of aerosols in characterizing the global solar radiation variation. The significant trends in air pressure and precipitation were irregularly distributed in space, and most regions did not show a significant change in these two variables. Besides evaluating the climate variables derived from GLDAS 2.1, we also assessed the performance of reconstructing the MODIS LAI product in 24 typical regions, which covered major climate and vegetation types, and also the global scale. The results showed that most of the outliers were detected and removed, and the fitted double logistics curves well characterized the variations and patterns of the annual LAI, reasonably caught the timing of vegetation phenology between the growing season and non-growing season, and retained the duration of the peak within the growing season for both single vegetation cycle and double vegetation cycles with average goodness of fit R^2 of 0.89 and 0.90 respectively.

In chapter 4, an empirical radiation partitioning approach was evaluated against 258 site-years from 36 flux tower sites, and the results showed that the modeled diffuse radiation was in general agreement with the observed diffuse radiation for different vegetation types with an average Pearson's correlation coefficient of 0.87. This radiation partitioning approach used in this study could help to derive the two radiation components: direct and diffuse when only the information of total solar radiation was available. We further evaluated the performance of the integrated two stream radiation transfer model that considered multiple scattering compared to Beer's law model in estimating solar radiation absorbed by vegetation canopy, since Beer's law probably overestimated the absorbed radiation due to a lack of considering scattering. The absorption fraction simulated by the two stream approach used in this study was lower than that estimated by Beer's law regardless of the LAI and diffuse radiation fraction, and the discrepancy in absorption fraction reached up to 73% in an overcast day. We further compared the simulated GPP derived using Beer's law model, two stream approach with the big-leaf model, and our integrated radiative transfer model (two stream approach with two-leaf model) with the GPP measurements obtained from six flux tower sites representing six vegetation types. The TS-TL model reduced the RMSE and bias by up to 72% and 81% based on the BL model, and up to 63% and 75% based on the TS-BL model, respectively. Moreover, the considerable improvements at larger LAI values by using the TS-TL model compared to the other two models indicated that considering multiple scattering or differentiating sunlit and shaded leaves largely reduced the biases in modeling the light profiles and estimating total photosynthesis by vegetation. Overall, our integrated RTM (TS-TL model) exhibited relatively robust and consistent performance in accurately simulating carbon assimilation for different LAI levels.

In chapter 5, simulated GPP based on the comprehensive process-based model was evaluated against 167 flux tower sites that covers 12 major vegetation types at both half-hour and annual scale. Our integrated model performed well in reproducing temporal variation in GPP at most flux tower sites, and the modeled GPP estimates were highly correlated to the flux tower observations for all the vegetation types ($r > 0.75$) at a half-hourly scale. Annually, our simulated GPP showed a strong significant linear relationship with the flux tower data, and it explained 72% of the spatial variations in GPP across all the validation sites. The positive bias showed a slight overestimation in annual GPP. In addition, global terrestrial GPP at 0.25° spatial resolution and 3-hourly temporal resolution was simulated using our integrated process-based model from 2001 to 2020 driven by reconstructed LAI data from MODIS MCD15A3H and climate data from GLDAS 2.1. The annual global terrestrial GPP ranged from 118 PgC to 134 PgC, with an average of 128 PgC, during the study period, and it showed a significantly increasing trend with an average rate of 0.71 PgC/yr globally. The annual GPP also varied in different vegetation types, and significant increasing was found in all the vegetation types except in evergreen broadleaf forests. The spatial patterns of GPP were not homogenous in space, and approximately 83.8% of global terrestrial showed increasing trend in annual GPP during 2001-2020, where the significantly increasing trends were mainly distributed in temperate regions and high elevation regions. Comparing to several recent GPP estimates and products, our simulated results were within a reasonable range of global terrestrial GPP estimations but with some discrepancies due to the different models, parameters, and driven data used to simulate GPP. In addition, the sensitivity analysis exhibited that our simulated GPP was most sensitive to biophysiological parameters V_{cmax25} and LAI, emphasizing the needs of obtaining accurate biophysiological parameters in large scale.

6.2. Uncertainties and limitations

Even though the methods and models used in this study were shown to perform relatively well and were robust in estimating global terrestrial GPP of various vegetation types, there still exists uncertainties and limitations that need further improvements. First, climate variables are essential in driving ecosystem models to estimate global terrestrial GPP, and obtaining accurate and consistent time series at a global scale becomes vital in global carbon modeling. Although GLDAS 2.1 showed reliable estimates in most variables, large biases occurred in precipitation which is one of the key variables that affect ecosystem models. Evaluations of different climate datasets could help to improve the accuracy of the input driven data in ecosystem models. In addition, the spatial resolution of 0.25 degree and temporal resolution of 3-hour when simulating global terrestrial GPP in this study was limited by the resolution of climate variables. A climate dataset with finer resolution could improve the GPP simulation accuracy and provide more information on spatial and temporal variations. Then, an empirical radiation partitioning model was used to obtain the direct and diffuse components from total solar radiation. Although the radiation partitioning approach generally captured the patterns of partitioning, the negative biases between our estimated data and observations exhibited an overall tendency of slightly underestimation in diffuse radiation. Since diffuse radiation is generated by the scattering effects of molecules and aerosols in atmosphere and is directly affected by clouds and aerosols, taking the aerosol factors into consideration could better simulate the direct and diffuse radiation. Also, leaf structures and foliar clumping effects should be considered in the radiatve transfer model as they influence the proportion of sunlit and shaded leaves and consequently change carbon assimilations by vegetation canopy.

Appendices

Table A.1 Information of flux tower stations used for evaluating the climate variables derived from GLDAS 2.1.

Vegetation Type	Site ID	Site name	Latitude (°)	Longitude (°)	Time Period
CRO	BE-Lon	Lonzee	50.55	4.75	2004-2014
	DE-Geb	Gebesee	51.10	10.91	2001-2014
	DE-Kli	Klingenberg	50.89	13.52	2004-2014
	DE-RuS	Selhausen Juelich	50.87	6.45	2011-2014
	DE-Seh	Selhausen	50.87	6.45	2007-2010
	DK-Fou	Foulum	56.48	9.59	2005-2005
	FI-Jok	Jokioinen	60.90	23.51	2001-2003
	FR-Gri	Grignon	48.84	1.95	2004-2014
	US-ARM	Southern Great Plains site Lamont	36.61	-97.49	2003-2012
	US-CRT	Curtice Walter-Berger cropland	41.63	-83.35	2011-2013
	US-Ne1	Mead-irrigated continuous maize site	41.17	-96.48	2001-2013
	US-Ne2	Mead-irrigated maize-soybean rotation site	41.16	-96.47	2001-2013
	US-Ne3	Mead-rainfed maize-soybean rotation site	41.18	-96.44	2001-2013
	US-Tw2	Twitchell Corn	38.10	-121.64	2012-2013
	US-Tw3	Twitchell Alfalfa	38.12	-121.65	2013-2014
CSH	RU-Vrk	Seida/Vorkuta	67.05	62.94	2008-2008
DBF	CA-Oas	Saskatchewan - Western Boreal, Mature Aspen	53.63	-106.20	2000-2010
	CA-TPD	Ontario - Turkey Point Mature Deciduous	42.64	-80.56	2012-2014
	DE-Hai	Hainich	51.08	10.45	2001-2012
	DE-Lnf	Leinefelde	51.33	10.37	2002-2012
	DK-Sor	Soroe	55.49	11.64	2000-2014
	FR-Fon	Fontainebleau-Barbeau	48.48	2.78	2005-2014
	IT-CA1	Castel d'Asso1	42.38	12.03	2011-2014

	IT-CA3	Castel d'Asso 3	42.38	12.02	2011-2014
	IT-PT1	Parco Ticino forest	45.20	9.06	2002-2004
	IT-Ro1	Roccarespampani 1	42.41	11.93	2001-2008
	IT-Ro2	Roccarespampani 2	42.39	11.92	2002-2012
	US-Ha1	Harvard Forest EMS Tower	42.54	-72.17	2000-2012
	US-Oho	Oak Openings	41.55	-83.84	2004-2013
	US-UMB	Univ. of Mich. Biological Station	45.56	-84.71	2001-2014
	US-UMd	UMBS Disturbance	45.56	-84.70	2007-2014
	US-WCr	Willow Creek	45.81	-90.08	2000-2014
	US-Wi1	Intermediate hardwood	46.73	-91.23	2003-2003
	US-Wi3	Mature hardwood	46.63	-91.10	2002-2004
	US-Wi8	Young hardwood clearcut	46.72	-91.25	2002-2002
DNF	RU-SkP	Yakutsk Spasskaya Pad larch	62.26	129.17	2012-2014
EBF	AU-Cum	Cumberland Plain	-33.62	150.72	2012-2014
	AU-Whr	Whroo	-36.67	145.03	2011-2014
	AU-Wom	Wombat	-37.42	144.09	2010-2014
	CN-Din	Dinghushan	23.17	112.54	2003-2005
	FR-Pue	Puechabon	43.74	3.60	2001-2014
ENF	AU-ASM	Alice Springs	-22.28	133.25	2010-2014
	CA-Qfo	Quebec - Eastern Boreal, Mature Black Spruce	49.69	-74.34	2003-2010
	CA-SF1	Saskatchewan - Western Boreal, forest burned in 1977	54.49	-105.82	2003-2006
	CA-SF2	Saskatchewan - Western Boreal, forest burned in 1989	54.25	-105.88	2001-2005
	CA-TP1	Ontario - Turkey Point 2002 Plantation White Pine	42.66	-80.56	2002-2014
	CA-TP2	Ontario - Turkey Point 1989 Plantation White Pine	42.77	-80.46	2002-2007
	CN-Qia	Qianyanzhou	26.74	115.06	2003-2005
	FI-Hyy	Hyytiala	61.85	24.29	2000-2014
	FI-Let	Lettosuo	60.64	23.96	2009-2012
	FI-Sod	Sodankyla	67.36	26.64	2001-2014
	FR-LBr	Le Bray	44.72	-0.77	2000-2008

	IT-La2	Lavarone2	45.95	11.29	2001-2002
	IT-SRo	San Rossore	43.73	10.28	2000-2012
	NL-Loo	Loobos	52.17	5.74	2000-2014
	RU-Fyo	Fyodorovskoye	56.46	32.92	2000-2014
	US-Me2	Metolius mature ponderosa pine	44.45	-121.56	2002-2014
	US-Wi2	Intermediate red pine	46.69	-91.15	2003-2003
	US-Wi4	Mature red pine	46.74	-91.17	2002-2005
	US-Wi5	Mixed young jack pine	46.65	-91.09	2004-2004
	US-Wi9	Young Jack pine	46.62	-91.08	2004-2005
GRA	AU-Emr	Emerald	-23.86	148.47	2011-2013
	AU-Rig	Riggs Creek	-36.65	145.58	2011-2014
	AU-Stp	Sturt Plains	-17.15	133.35	2008-2014
	AU-Ync	Jaxa	-34.99	146.29	2012-2014
	CH-Cha	Chamau	47.21	8.41	2005-2014
	CH-Oe1	Oensingen grassland	47.29	7.73	2002-2008
	CN-Cng	Changling	44.59	123.51	2007-2010
	CN-Du2	Duolun_grassland	42.05	116.28	2006-2008
	CN-Du3	Duolun Degraded Meadow	42.06	116.28	2009-2010
	CN-Sw2	Siziwang Grazed	41.79	111.90	2010-2012
	DE-Gri	Grillenburg	50.95	13.51	2004-2014
	DE-RuR	Rollesbroich	50.62	6.30	2011-2014
	DK-Eng	Enghave	55.69	12.19	2005-2008
	DK-ZaH	Zackenberg Heath	74.47	-20.55	2001-2014
	NL-Hor	Horstermeer	52.24	5.07	2004-2011
	RU-Ha1	Hakasia steppe	54.73	90.00	2002-2004
	RU-Sam	Samoylov	72.37	126.50	2002-2014
	RU-Tks	Tiksi	71.59	128.89	2010-2014
	US-AR1	ARM USDA UNL OSU Woodward Switchgrass 1	36.43	-99.42	2009-2012
	US-AR2	ARM USDA UNL OSU Woodward Switchgrass 2	36.64	-99.60	2009-2012
	US-ARb	ARM Southern Great Plains burn site- Lamont	35.55	-98.04	2005-2006

	US-ARc	ARM Southern Great Plains control site- Lamont	35.55	-98.04	2005-2006
	US-Goo	Goodwin Creek	34.25	-89.87	2002-2006
	US-IB2	Fermi National Accelerator Laboratory- Batavia	41.84	-88.24	2004-2011
	US-SRG	Santa Rita Grassland	31.79	-110.83	2008-2014
	US-Var	Vaira Ranch- Ione	38.41	-120.95	2001-2014
	US-Wkg	Walnut Gulch Kendall Grasslands	31.74	-109.94	2004-2014
MF	BE-Bra	Brasschaat	51.31	4.52	2000-2014
	BE-Vie	Vielsalm	50.31	6.00	2000-2014
	CA-Gro	Ontario - Groundhog River, Boreal Mixedwood Forest	48.22	-82.16	2003-2014
	CH-Lae	Laegern	47.48	8.37	2004-2014
	CN-Cha	Changbaishan	42.40	128.10	2003-2005
	JP-SMF	Seto Mixed Forest Site	35.26	137.08	2002-2006
	US-Syv	Sylvania Wilderness Area	46.24	-89.35	2001-2014
OSH	CA-NS6	UCI-1989 burn site	55.92	-98.96	2001-2005
	CA-NS7	UCI-1998 burn site	56.64	-99.95	2002-2005
	CA-SF3	Saskatchewan - Western Boreal, forest burned in 1998	54.09	-106.01	2001-2006
	RU-Cok	Chokurdakh	70.83	147.49	2003-2014
	US-Whs	Walnut Gulch Lucky Hills Shrub	31.74	-110.05	2007-2014
SAV	AU-Cpr	Calperum	-34.00	140.59	2010-2014
	AU-DaS	Daly River Cleared	-14.16	131.39	2008-2014
	AU-Dry	Dry River	-15.26	132.37	2008-2014
	AU-GWW	Great Western Woodlands, Western Australia, Australia	-30.19	120.65	2013-2014
SNO	NO-Blv	Bayelva, Spitsbergen	78.92	11.83	2008-2009
WET	CZ-wet	Trebon	49.02	14.77	2006-2014
	DE-Akm	Anklam	53.87	13.68	2009-2014
	DE-Zrk	Zarnekow	53.88	12.89	2013-2014
	DK-NuF	Nuuk Fen	64.13	-51.39	2008-2014
	DK-ZaF	Zackenberg Fen	74.48	-20.55	2008-2011
	FI-Lom	Lompolojankka	68.00	24.21	2007-2009

	US-Atq	Atqasuk	70.47	-157.41	2003-2008
	US-Los	Lost Creek	46.08	-89.98	2001-2014
	US-Tw1	Twitchell Wetland West Pond	38.11	-121.65	2012-2014
	US-Tw4	Twitchell East End Wetland	38.10	-121.64	2013-2014
WSA	AU-Gin	Gingin	-31.38	115.71	2011-2014
	AU-How	Howard Springs	-12.49	131.15	2001-2014
	AU-RDF	Red Dirt Melon Farm, Northern Territory	-14.56	132.48	2011-2013
	US-SRM	Santa Rita Mesquite	31.82	-110.87	2004-2014
	US-Ton	Tonzi Ranch	38.43	-120.97	2001-2014

Table A.2 Information of flux tower stations used for evaluating radiation partitioning.

Vegetation Type	Site ID	Site name	Latitude (°)	Longitude (°)	Time Period
CRO	CH-Oe2	Oensingen crop	47.29	7.73	2004-2012
	DE-Geb	Gebesee	51.10	10.91	2003-2014
	FR-Gri	Grignon	48.84	1.95	2004-2014
	IT-BCi	Borgo Cioffi	40.52	14.96	2005-2011
DBF	CA-Oas	Saskatchewan - Western Boreal, Mature Aspen	53.63	-106.20	2003-2010
	DE-Hai	Hainich	51.08	10.45	2003-2012
	DE-Lnf	Leinefelde	51.33	10.37	2003-2012
	DK-Sor	Soroe	55.49	11.64	2003-2013
	IT-Col	Collelongo	41.85	13.59	2004-2014
	IT-PT1	Parco Ticino forest	45.20	9.06	2003-2004
	IT-Ro2	Roccarespampani 2	42.39	11.92	2004-2004
	US-UMd	UMBS Disturbance	45.56	-84.70	2008-2010
EBF	FR-Pue	Puechabon	43.74	3.60	2003-2014
	GF-Guy	Guyaflux (French Guiana)	5.28	-52.92	2008-2014
	GH-Ank	Ankasa	5.27	-2.69	2011-2014
ENF	CA-Obs	Saskatchewan - Western Boreal, Mature Black Spruce	53.99	-105.12	2003-2010
	CA-Qfo	Quebec - Eastern Boreal, Mature Black Spruce	49.69	-74.34	2003-2010
	CZ-BK1	Bily Kriz forest	49.50	18.54	2004-2014
	DE-Tha	Tharandt	50.96	13.57	2003-2014
	FI-Hyy	Hyytiala	61.85	24.29	2003-2014
	FR-LBr	Le Bray	44.72	-0.77	2003-2008
	IT-Ren	Renon	46.59	11.43	2003-2013
	IT-SR2	San Rossore 2	43.73	10.29	2013-2014
	IT-SRo	San Rossore	43.73	10.28	2004-2012
	NL-Loo	Loobos	52.17	5.74	2004-2014
	US-Me1	Metolius - Eyerly burn	44.58	-121.50	2004-2005
	US-Me2	Metolius mature ponderosa pine	44.45	-121.56	2012-2014
GRA	AT-Neu	Neustift	47.12	11.32	2003-2012

	CH-Oe1	Oensingen grassland	47.29	7.73	2008-2008
	CZ-BK2	Bily Kriz grassland	49.49	18.54	2004-2006
	IT-MBo	Monte Bondone	46.01	11.05	2003-2013
	RU-Ha1	Hakasia steppe	54.73	90.00	2003-2004
MF	CA-Gro	Ontario - Groundhog River, Boreal Mixedwood Forest	48.22	-82.16	2003-2014
	US-Syv	Sylvania Wilderness Area	46.24	-89.35	2003-2003
OSH	ES-Ln2	Lanjaron-Salvage logging	36.97	-3.48	2009-2009
SAV	CG-Tch	Tchizalamou	-4.29	11.66	2006-2009

Table A.3 Information of flux tower sites used for GPP validation.

Vegetation Type	Site ID	Site name	Latitude (°)	Longitude (°)	Time Period
CRO	BE-Lon	Lonzee	50.55	4.75	2004-2014
	CH-Oe2	Oensingen crop	47.29	7.73	2004-2014
	DE-Geb	Gebesee	51.10	10.91	2003-2014
	DE-Kli	Klingenberg	50.89	13.52	2004-2014
	DE-Seh	Selhausen	50.87	6.45	2007-2010
	DK-Fou	Foulum	56.48	9.59	2005-2005
	FI-Jok	Jokioinen	60.90	23.51	2003-2003
	FR-Gri	Grignon	48.84	1.95	2004-2014
	IT-BCi	Borgo Cioffi	40.52	14.96	2004-2014
	IT-CA2	Castel d'Asso2	42.38	12.03	2011-2014
	US-ARM	Southern Great Plains site- Lamont	36.61	-97.49	2003-2012
	US-CRT	Curtice Walter-Berger cropland	41.63	-83.35	2011-2013
	US-Lin	Lindcove Orange Orchard	36.36	-119.84	2009-2010
	US-Tw2	Twitchell Corn	38.10	-121.64	2012-2013
	US-Tw3	Twitchell Alfalfa	38.12	-121.65	2013-2014
	US-Twt	Twitchell Island	38.11	-121.65	2009-2014
CSH	IT-Noe	Arca di Noe - Le Prigionette	40.61	8.15	2004-2014
	US-KS2	Kennedy Space Center (scrub oak)	28.61	-80.67	2003-2006
DBF	AU-Lox	Loxton	-34.47	140.66	2008-2009
	CA-Oas	Saskatchewan	53.63	-106.20	2003-2010
	CA-TPD	Turkey Point Mature Deciduous	42.64	-80.56	2012-2014
	DE-Hai	Hainich	51.08	10.45	2003-2012
	DE-Lnf	Leinefelde	51.33	10.37	2003-2012
	DK-Sor	Soroe	55.49	11.64	2003-2014
	FR-Fon	Fontainebleau-Barbeau	48.48	2.78	2005-2014
	IT-CA1	Castel d'Asso1	42.38	12.03	2011-2014
	IT-CA3	Castel d'Asso 3	42.38	12.02	2011-2014
	IT-Col	Collelongo	41.85	13.59	2003-2014
	IT-Isp	Ispra ABC-IS	45.81	8.63	2013-2014
	IT-PT1	Parco Ticino forest	45.20	9.06	2003-2004
	IT-Ro2	Roccarespampani 2	42.39	11.92	2003-2012

	JP-MBF	Moshiri Birch Forest Site	44.39	142.32	2004-2005
	PA-SPn	Sardinilla Plantation	9.32	-79.63	2007-2009
	US-Oho	Oak Openings	41.55	-83.84	2004-2013
	US-UMd	UMBS Disturbance	45.56	-84.70	2007-2014
	US-WCr	Willow Creek	45.81	-90.08	2003-2014
	US-Wi1	Intermediate hardwood (IHW)	46.73	-91.23	2003-2003
	US-Wi3	Mature hardwood (MHW)	46.63	-91.10	2004-2004
	ZM-Mon	Mongu	-15.44	23.25	2007-2009
DNF	RU-SkP	Yakutsk Spasskaya Pad larch	62.26	129.17	2012-2014
EBF	AU-Cum	Cumberland Plain	-33.62	150.72	2012-2014
	AU-Rob	Robson Creek, Queensland	-17.12	145.63	2014-2014
	AU-Wac	Wallaby Creek	-37.43	145.19	2005-2008
	AU-Whr	Whroo	-36.67	145.03	2011-2014
	AU-Wom	Wombat	-37.42	144.09	2010-2014
	BR-Sa3	Santarem-Km83-Logged Forest	-3.02	-54.97	2003-2004
	CN-Din	Dinghushan	23.17	112.54	2003-2005
	FR-Pue	Puechabon	43.74	3.60	2003-2014
	GF-Guy	Guyaflux (French Guiana)	5.28	-52.92	2004-2014
	GH-Ank	Ankasa	5.27	-2.69	2011-2014
	IT-Cp2	Castelporziano2	41.70	12.36	2012-2014
	IT-Cpz	Castelporziano	41.71	12.38	2003-2008
	MY-PSO	Pasoh Forest Reserve (PSO)	2.97	102.31	2003-2009
ENF	AR-Vir	Virasoro	-28.24	-56.19	2010-2012
	AU-ASM	Alice Springs	-22.28	133.25	2010-2014
	CA-NS1	UCI-1850 burn site	55.88	-98.48	2003-2005
	CA-NS2	UCI-1930 burn site	55.91	-98.52	2003-2005
	CA-NS3	UCI-1964 burn site	55.91	-98.38	2003-2005
	CA-NS4	UCI-1964 burn site wet	55.91	-98.38	2003-2005
	CA-NS5	UCI-1981 burn site	55.86	-98.49	2003-2005
	CA-Obs	Saskatchewan Mature Black Spruce	53.99	-105.12	2003-2010
	CA-Qfo	Quebec - Mature Black Spruce	49.69	-74.34	2003-2010
	CA-SF1	Saskatchewan forest burned in 1977	54.49	-105.82	2003-2006
	CA-SF2	Saskatchewan forest burned in 1989	54.25	-105.88	2003-2005
	CA-TP1	Turkey Point 2002	42.66	-80.56	2003-2014

	CA-TP2	Turkey Point 1989	42.77	-80.46	2003-2007
	CA-TP3	Turkey Point 1974	42.71	-80.35	2003-2014
	CA-TP4	Turkey Point 1939	42.71	-80.36	2003-2014
	CH-Dav	Davos	46.82	9.86	2003-2014
	CN-Qia	Qianyanzhou	26.74	115.06	2003-2005
	CZ-BK1	Bily Kriz forest	49.50	18.54	2004-2014
	DE-Lkb	Lackenberg	49.10	13.30	2009-2013
	DE-Obe	Oberbärenburg	50.79	13.72	2008-2014
	DE-Tha	Tharandt	50.96	13.57	2003-2014
	FI-Hyy	Hyytiala	61.85	24.29	2003-2014
	FI-Let	Lettosuo	60.64	23.96	2009-2012
	FR-LBr	Le Bray	44.72	-0.77	2003-2008
	IT-Lav	Lavarone	45.96	11.28	2003-2014
	IT-Ren	Renon	46.59	11.43	2003-2013
	IT-SR2	San Rossore 2	43.73	10.29	2013-2014
	IT-SRo	San Rossore	43.73	10.28	2003-2012
	NL-Loo	Loobos	52.17	5.74	2003-2014
	RU-Fyo	Fyodorovskoye	56.46	32.92	2003-2014
	US-Blo	Blodgett Forest	38.90	-120.63	2003-2007
	US-GBT	GLEES Brooklyn Tower	41.37	-106.24	2003-2003
	US-GLE	GLEES	41.37	-106.24	2005-2014
	US-Me1	Metolius - Eyerly burn	44.58	-121.50	2004-2005
	US-Me2	Metolius mature ponderosa pine	44.45	-121.56	2003-2014
	US-Me3	Metolius-second young aged pine	44.32	-121.61	2004-2009
	US-Me6	Metolius Young Pine Burn	44.32	-121.61	2010-2014
	US-NR1	Niwot Ridge Forest (LTER NWT1)	40.03	-105.55	2003-2014
	US-Wi4	Mature red pine (MRP)	46.74	-91.17	2003-2005
	US-Wi5	Mixed young jack pine (MYJP)	46.65	-91.09	2004-2004
	US-Wi9	Young Jack pine (YJP)	46.62	-91.08	2004-2005
GRA	AT-Neu	Neustift	47.12	11.32	2003-2012
	AU-DaP	Daly River Savanna	-14.06	131.32	2007-2013
	AU-Emr	Emerald	-23.86	148.47	2011-2013
	AU-Rig	Riggs Creek	-36.65	145.58	2011-2014
	AU-Stp	Sturt Plains	-17.15	133.35	2008-2014

	CH-Cha	Chamau	47.21	8.41	2005-2014
	CH-Fru	Früebüel	47.12	8.54	2005-2014
	CH-Oe1	Oensingen grassland	47.29	7.73	2003-2008
	CN-Cng	Changling	44.59	123.51	2007-2010
	CN-Du2	Duolun_grassland (D01)	42.05	116.28	2007-2008
	CN-Du3	Duolun Degraded Meadow	42.06	116.28	2009-2010
	CN-HaM	Haibei Alpine Tibet site	37.37	101.18	2003-2004
	CZ-BK2	Bily Kriz grassland	49.49	18.54	2006-2012
	DE-Gri	Grillenburg	50.95	13.51	2004-2014
	DE-RuR	Rollesbroich	50.62	6.30	2011-2014
	DK-Eng	Enghave	55.69	12.19	2005-2008
	IT-MBo	Monte Bondone	46.01	11.05	2003-2013
	IT-Tor	Torgnon	45.84	7.58	2008-2014
	NL-Hor	Horstermeer	52.24	5.07	2004-2011
	PA-SPs	Sardinilla-Pasture	9.31	-79.63	2007-2009
	RU-Ha1	Hakasia steppe	54.73	90.00	2003-2004
	US-AR1	ARM USDA UNL OSU Woodward Switchgrass 1	36.43	-99.42	2009-2012
	US-AR2	ARM USDA UNL OSU Woodward Switchgrass 2	36.64	-99.60	2009-2012
	US-ARb	ARM Southern Great Plains burn site- Lamont	35.55	-98.04	2005-2006
	US-ARc	ARM Southern Great Plains control site- Lamont	35.55	-98.04	2005-2006
	US-Goo	Goodwin Creek	34.25	-89.87	2003-2006
	US-IB2	Fermi National Accelerator Laboratory- Batavia (Prairie site)	41.84	-88.24	2004-2011
	US-SRG	Santa Rita Grassland	31.79	-110.83	2008-2014
	US-Var	Vaira Ranch- Ione	38.41	-120.95	2003-2014
	US-Wkg	Walnut Gulch Kendall Grasslands	31.74	-109.94	2004-2014
MF	AR-SLu	San Luis	-33.46	-66.46	2009-2011
	BE-Bra	Brasschaat	51.31	4.52	2004-2014
	BE-Vie	Vielsalm	50.31	6.00	2003-2014
	CA-Gro	Ontario	48.22	-82.16	2003-2014

	CH-Lae	Laegern	47.48	8.37	2004-2014
	CN-Cha	Changbaishan	42.40	128.10	2003-2005
	JP-SMF	Seto Mixed Forest Site	35.26	137.08	2003-2006
	US-Syv	Sylvania Wilderness Area	46.24	-89.35	2003-2014
OSH	AU-TTE	Ti Tree East	-22.29	133.64	2012-2014
	CA-NS6	UCI-1989 burn site	55.92	-98.96	2003-2005
	CA-NS7	UCI-1998 burn site	56.64	-99.95	2003-2005
	CA-SF3	Saskatchewan - Western Boreal, forest burned in 1998	54.09	-106.01	2003-2006
	ES-LgS	Laguna Seca	37.10	-2.97	2007-2009
	ES-LJu	Llano de los Juanes	36.93	-2.75	2004-2013
	US-SRC	Santa Rita Creosote	31.91	-110.84	2008-2014
	US-Sta	Saratoga	41.40	-106.80	2005-2009
	US-Whs	Walnut Gulch Lucky Hills Shrub	31.74	-110.05	2007-2014
	US-Wi6	Pine barrens #1 (PB1)	46.62	-91.30	2003-2003
	US-Wi7	Red pine clearcut (RPCC)	46.65	-91.07	2005-2005
SAV	AU-Cpr	Calperum	-34.00	140.59	2010-2014
	AU-DaS	Daly River Cleared	-14.16	131.39	2008-2014
	AU-Dry	Dry River	-15.26	132.37	2008-2014
	AU-GWW	Great Western Woodlands, Western Australia	-30.19	120.65	2013-2014
	CG-Tch	Tchizalamou	-4.29	11.66	2006-2009
	SD-Dem	Demokeya	13.28	30.48	2005-2009
	SN-Dhr	Dahra	15.40	-15.43	2010-2013
	ZA-Kru	Skukuza	-25.02	31.50	2003-2013
WET	AU-Fog	Fogg Dam	-12.55	131.31	2006-2008
	CN-Ha2	Haibei Shrubland	37.61	101.33	2003-2005
	CZ-wet	Trebon (CZECHWET)	49.02	14.77	2006-2014
	DE-SfN	Schechenfilz Nord	47.81	11.33	2012-2014
	DE-Spw	Spreewald	51.89	14.03	2010-2014
	DE-Zrk	Zarnekow	53.88	12.89	2013-2014
	US-Los	Lost Creek	46.08	-89.98	2003-2014
	US-Myb	Mayberry Wetland	38.05	-121.77	2011-2014
	US-Tw1	Twitchell Wetland West Pond	38.11	-121.65	2012-2014

	US-Tw4	Twitchell East End Wetland	38.10	-121.64	2013-2014
WSA	AU-Ade	Adelaide River	-13.08	131.12	2007-2009
	AU-Gin	Gingin	-31.38	115.71	2011-2014
	AU-How	Howard Springs	-12.49	131.15	2003-2014
	AU-RDF	Red Dirt Melon Farm	-14.56	132.48	2011-2013
	US-SRM	Santa Rita Mesquite	31.82	-110.87	2004-2014
	US-Ton	Tonzi Ranch	38.43	-120.97	2003-2014

References

Alton, P.B., North, P.R. and Los, S.O., 2007. The impact of diffuse sunlight on canopy light-use efficiency, gross photosynthetic product and net ecosystem exchange in three forest biomes. Global Change Biology, 13(4), pp.776-787.

Amiro, B.D., Barr, A.G., Black, T.A., Iwashita, H., Kljun, N., McCaughey, J.H., Morgenstern, K., Murayama, S., Nesic, Z., Orchansky, A.L. and Saigusa, N., 2006. Carbon, energy and water fluxes at mature and disturbed forest sites, Saskatchewan, Canada. Agricultural and forest meteorology, 136(3-4), pp.237-251.

Amiro, B.D., Barr, A.G., Barr, J.G., Black, T.A., Bracho, R., Brown, M., Chen, J., Clark, K.L., Davis, K.J., Desai, A.R. and Dore, S., 2010. Ecosystem carbon dioxide fluxes after disturbance in forests of North America. Journal of Geophysical Research: Biogeosciences, 115(G4).

Amthor, J.S., 1994. Scaling CO 2-photosynthesis relationships from the leaf to the canopy. Photosynthesis Research, 39(3), pp.321-350.

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., Murray-Tortarolo, G., Papale, D., Parazoo, N.C., Peylin, P. and Piao, S., 2015. Spatiotemporal patterns of terrestrial gross primary production: A review. Reviews of Geophysics, 53(3), pp.785-818.

Baldocchi, D.D., Hincks, B.B. and Meyers, T.P., 1988. Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods. Ecology, 69(5), pp.1331-1340.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R. and Fuentes, J., 2001. FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bulletin of the American Meteorological Society, 82(11), pp.2415-2434.

Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Niño, F., Weiss, M., Samain, O. and Roujean, J.L., 2007. LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION: Part 1: Principles of the algorithm. Remote sensing of environment, 110(3), pp.275-286.

Bernacchi, C.J., Singsaas, E.L., Pimentel, C., Portis Jr, A.R. and Long, S.P., 2001. Improved temperature response functions for models of Rubisco-limited photosynthesis. Plant, Cell & Environment, 24(2), pp.253-259.

Bi, W., He, W., Zhou, Y., Ju, W., Liu, Y., Liu, Y., Zhang, X., Wei, X. and Cheng, N., 2022. A global 0.05° dataset for gross primary production of sunlit and shaded vegetation canopies from 1992 to 2020. Scientific Data, 9(1), pp.1-13.

Bonan, G.B., Lawrence, P.J., Oleson, K.W., Levis, S., Jung, M., Reichstein, M., Lawrence, D.M. and Swenson, S.C., 2011. Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. Journal of Geophysical Research: Biogeosciences, 116(G2).

Bosilovich, M.G., Chen, J., Robertson, F.R. and Adler, R.F., 2008. Evaluation of global precipitation in reanalyses. Journal of applied meteorology and climatology, 47(9), pp.2279-2299.

Cai, Z., Jönsson, P., Jin, H. and Eklundh, L., 2017. Performance of smoothing methods for reconstructing NDVI time-series and estimating vegetation phenology from MODIS data. Remote Sensing, 9(12), p.1271.

Collatz, G.J., Ribas-Carbo, M. and Berry, J.A., 1992. Coupled photosynthesis-stomatal

conductance model for leaves of C4 plants. Functional Plant Biology, 19(5), pp.519-538.
Cramer, W., Bondeau, A., Woodward, F.I., Prentice, I.C., Betts, R.A., Brovkin, V., Cox, P.M., Fisher, V., Foley, J.A., Friend, A.D. and Kucharik, C., 2001. Global response of terrestrial ecosystem structure and function to CO2 and climate change: results from six dynamic global vegetation models. Global change biology, 7(4), pp.357-373.
Dai, Y., Dickinson, R.E. and Wang, Y.P., 2004. A two-big-leaf model for canopy temperature, photosynthesis, and stomatal conductance. Journal of Climate, 17(12), pp.2281-2299.
De Pury, D.G.G. and Farquhar, G.D., 1997. Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models. Plant, Cell & Environment, 20(5), pp.537-557.
Decker, M., Brunke, M.A., Wang, Z., Sakaguchi, K., Zeng, X. and Bosilovich, M.G., 2012. Evaluation of the reanalysis products from GSFC, NCEP, and ECMWF using flux tower observations. Journal of Climate, 25(6), pp.1916-1944.
Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, D.P. and Bechtold, P., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quarterly Journal of the royal meteorological society, 137(656), pp.553-597.
Deng, F., Chen, J.M., Plummer, S., Chen, M. and Pisek, J., 2006. Algorithm for global leaf area index retrieval using satellite imagery. IEEE transactions on geoscience and remote sensing, 44(8), pp.2219-2229.
Dickinson, R.E., 1983. Land surface processes and climate—Surface albedos and energy balance. In Advances in geophysics (Vol. 25, pp. 305-353). Elsevier.
Dixon, R.K. and Turner, D.P., 1991. The global carbon cycle and climate change: responses and feedbacks from below-ground systems. Environmental pollution, 73(3-4), pp.245-262.
Ezhova, E., Ylivinkka, I., Kuusk, J., Komsaare, K., Vana, M., Krasnova, A., Noe, S., Arshinov, M., Belan, B., Park, S.B. and Lavrič, J.V., 2018. Direct effect of aerosols on solar radiation and gross primary production in boreal and hemiboreal forests. Atmospheric Chemistry and Physics, 18(24), pp.17863-17881.
Farquhar, G.D., von Caemmerer, S.V. and Berry, J.A., 1980. A biochemical model of photosynthetic CO 2 assimilation in leaves of C 3 species. Planta, 149(1), pp.78-90.
Friend, A.D., Arneth, A., Kiang, N.Y., Lomas, M., Ogee, J., Rödenbeck, C., Running, S.W., SANTAREN, J.D., Sitch, S., Viovy, N. and Ian Woodward, F., 2007. FLUXNET and modelling the global carbon cycle. Global Change Biology, 13(3), pp.610-633.
Gonsamo, A., Chen, J.M., Wu, C. and Dragoni, D., 2012. Predicting deciduous forest carbon uptake phenology by upscaling FLUXNET measurements using remote sensing data. Agricultural and forest meteorology, 165, pp.127-135.
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. and Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote sensing of Environment, 202, pp.18-27.
Gu, L., Baldocchi, D., Verma, S.B., Black, T.A., Vesala, T., Falge, E.M. and Dowty, P.R., 2002. Advantages of diffuse radiation for terrestrial ecosystem productivity. Journal of Geophysical Research: Atmospheres, 107(D6), pp.ACL-2.
Guan, X., Chen, J.M., Shen, H., Xie, X. and Tan, J., 2022. Comparison of big-leaf and two-leaf light use efficiency models for GPP simulation after considering a radiation scalar. Agricultural and Forest Meteorology, 313, p.108761.
Howell, T.A. and Dusek, D.A., 1995. Comparison of vapor-pressure-deficit calculation

methods—southern high plains. Journal of irrigation and drainage engineering, 121(2), pp.191-198.

Jarvis, P.G., 1976. The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 273(927), pp.593-610.

Ji, L., Senay, G.B. and Verdin, J.P., 2015. Evaluation of the Global Land Data Assimilation System (GLDAS) air temperature data products. Journal of Hydrometeorology, 16(6), pp.2463-2480.

Jiang, C. and Ryu, Y., 2016. Multi-scale evaluation of global gross primary productivity and evapotranspiration products derived from Breathing Earth System Simulator (BESS). Remote Sensing of Environment, 186, pp.528-547.

Kala, J., Decker, M., Exbrayat, J.F., Pitman, A.J., Carouge, C., Evans, J.P., Abramowitz, G. and Mocko, D., 2014. Influence of leaf area index prescriptions on simulations of heat, moisture, and carbon fluxes. Journal of Hydrometeorology, 15(1), pp.489-503.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J. and Zhu, Y., 1996. The NCEP/NCAR 40-year reanalysis project. Bulletin of the American meteorological Society, 77(3), pp.437-472.

Kandasamy, S., Baret, F., Verger, A., Neveux, P. and Weiss, M., 2013. A comparison of methods for smoothing and gap filling time series of remote sensing observations-application to MODIS LAI products. Biogeosciences, 10(6), p.4055.

Kanniah, K.D., Beringer, J., North, P. and Hutley, L., 2012. Control of atmospheric particles on diffuse radiation and terrestrial plant productivity: A review. Progress in Physical Geography, 36(2), pp.209-237.

Keenan, T.F., Baker, I., Barr, A., Ciais, P., Davis, K., Dietze, M., Dragoni, D., Gough, C.M., Grant, R., Hollinger, D. and Hufkens, K., 2012. Terrestrial biosphere model performance for inter-annual variability of land-atmosphere CO2 exchange. Global Change Biology, 18(6), pp.1971-1987.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H. and Miyaoka, K., 2015. The JRA-55 reanalysis: General specifications and basic characteristics. Journal of the Meteorological Society of Japan. Ser. II, 93(1), pp.5-48.

Kondratyev, K.Y., Krapivin, V.F. and Varotsos, C., 2003. Global carbon cycle and climate change. Springer Science & Business Media.

Larsen, D.R. and Kershaw Jr, J.A., 1996. Influence of canopy structure assumptions on predictions from Beer's law. A comparison of deterministic and stochastic simulations. Agricultural and Forest Meteorology, 81(1-2), pp.61-77.

Lasslop, G., Reichstein, M., Papale, D., Richardson, A.D., Arneth, A., Barr, A., Stoy, P. and Wohlfahrt, G., 2010. Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. Global Change Biology, 16(1), pp.187-208.

Leuning, R., 1995. A critical appraisal of a combined stomatal-photosynthesis model for C3 plants. Plant, Cell & Environment, 18(4), pp.339-355.

Leuning, R., Kelliher, F.M., De Pury, D.G.G. and Schulze, E.D., 1995. Leaf nitrogen, photosynthesis, conductance and transpiration: scaling from leaves to canopies. Plant, Cell & Environment, 18(10), pp.1183-1200.

Li, F., Wang, X., Zhao, J., Zhang, X. and Zhao, Q., 2013. A method for estimating the gross primary production of alpine meadows using MODIS and climate data in China. International journal of remote sensing, 34(23), pp.8280-8300.

Ligot, G., Balandier, P., Courbaud, B. and Claessens, H., 2014. Forest radiative transfer models: which approach for which application?. Canadian journal of forest research, 44(5), pp.391-403.

Liu, Y., Liu, R. and Chen, J.M., 2012. Retrospective retrieval of long-term consistent global leaf area index (1981–2011) from combined AVHRR and MODIS data. Journal of Geophysical Research: Biogeosciences, 117(G4).

Liu, D., Cai, W., Xia, J., Dong, W., Zhou, G., Chen, Y., Zhang, H. and Yuan, W., 2014. Global validation of a process-based model on vegetation gross primary production using eddy covariance observations. PloS one, 9(11).

Liu, Y., Xiao, J., Ju, W., Zhu, G., Wu, X., Fan, W., Li, D. and Zhou, Y., 2018. Satellite-derived LAI products exhibit large discrepancies and can lead to substantial uncertainty in simulated carbon and water fluxes. Remote Sensing of Environment, 206, pp.174-188.

Liu, P., Barr, A.G., Zha, T., Black, T.A., Jassal, R.S., Nesic, Z., Helgason, W.D., Jia, X. and Tian, Y., 2022. Re-assessment of the climatic controls on the carbon and water fluxes of a boreal aspen forest over 1996-2016: changing sensitivity to long-term climatic conditions. Global Change Biology.

Lloyd, J., Grace, J., Miranda, A.C., Meir, P., Wong, S.C., Miranda, H.S., Wright, I.R., Gash, J.H.C. and McIntyre, J., 1995. A simple calibrated model of Amazon rainforest productivity based on leaf biochemical properties. Plant, Cell & Environment, 18(10), pp.1129-1145.

Los, S.O., Pollack, N.H., Parris, M.T., Collatz, G.J., Tucker, C.J., Sellers, P.J., Malmström, C.M., DeFries, R.S., Bounoua, L. and Dazlich, D.A., 2000. A global 9-yr biophysical land surface dataset from NOAA AVHRR data. Journal of Hydrometeorology, 1(2), pp.183-199.

Luo, X., Chen, J.M., Liu, J., Black, T.A., Croft, H., Staebler, R., He, L., Arain, M.A., Chen, B., Mo, G. and Gonsamo, A., 2018. Comparison of big-leaf, two-big-leaf, and two-leaf upscaling schemes for evapotranspiration estimation using coupled carbon-water modeling. Journal of Geophysical Research: Biogeosciences, 123(1), pp.207-225.

Ma, J., Yan, X., Dong, W. and Chou, J., 2015. Gross primary production of global forest ecosystems has been overestimated. Scientific reports, 5, p.10820.

Mahat, V. and Tarboton, D.G., 2012. Canopy radiation transmission for an energy balance snowmelt model. Water Resources Research, 48(1).

Masarie, K.A. and Tans, P.P., 1995. Extension and integration of atmospheric carbon dioxide data into a globally consistent measurement record. Journal of Geophysical Research: Atmospheres, 100(D6), pp.11593-11610.

Medlyn, B.E., Dreyer, E., Ellsworth, D., Forstreuter, M., Harley, P.C., Kirschbaum, M.U.F., Le Roux, X., Montpied, P., Strassemeyer, J., Walcroft, A. and Wang, K., 2002. Temperature response of parameters of a biochemically based model of photosynthesis. II. A review of experimental data. Plant, Cell & Environment, 25(9), pp.1167-1179.

Monsi, M. and Saeki, T., 1953. Uber den Lichtfaktor in den Pflanzen-gesellschaften und seine Bedeutung fur die Stoffproduktion. Jap. Journ. Bot., 14, pp.22-52.

Monteith, J.L., 1972. Solar radiation and productivity in tropical ecosystems. Journal of applied ecology, 9(3), pp.747-766.

Monteith, J. and Unsworth, M., 2013. Principles of environmental physics: plants, animals, and the atmosphere. Academic Press.

Mooney, P.A., Mulligan, F.J. and Fealy, R., 2011. Comparison of ERA-40, ERA-Interim and

NCEP/NCAR reanalysis data with observed surface air temperatures over Ireland. International Journal of Climatology, 31(4), pp.545-557.

Moorcroft, P.R., 2006. How close are we to a predictive science of the biosphere?. Trends in Ecology & Evolution, 21(7), pp.400-407.

Morales, P., Sykes, M.T., Prentice, I.C., Smith, P., Smith, B., Bugmann, H., Zierl, B., Friedlingstein, P., Viovy, N., Sabaté, S. and Sánchez, A., 2005. Comparing and evaluating process-based ecosystem model predictions of carbon and water fluxes in major European forest biomes. Global Change Biology, 11(12), pp.2211-2233.

Mutanga, O. and Kumar, L., 2019. Google Earth Engine Applications. Remote Sens. 2019, 11(5), 591; https://doi.org/10.3390/rs11050591.

Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.R. and Lotsch, A., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. Remote sensing of environment, 83(1-2), pp.214-231.

Nemani, R.R., Keeling, C.D., Hashimoto, H., Jolly, W.M., Piper, S.C., Tucker, C.J., Myneni, R.B. and Running, S.W., 2003. Climate-driven increases in global terrestrial net primary production from 1982 to 1999. science, 300(5625), pp.1560-1563.

Nijssen, B. and Lettenmaier, D.P., 1999. A simplified approach for predicting shortwave radiation transfer through boreal forest canopies. Journal of Geophysical Research: Atmospheres, 104(D22), pp.27859-27868.

Nilson, T. and Ross, J., 1997. Modeling radiative transfer through forest canopies: implications for canopy photosynthesis and remote sensing. In The use of remote sensing in the modeling of forest productivity (pp. 23-60). Springer, Dordrecht.

Norman J.M., 1980. Interfacing leaf and canopy irradiance inter-ception models. In Predieting Photosynthesis for EeosystemModels, Vol. II (eds J. D. Hesketh & J. W. Jones), pp. 49-67.CRC Press, Inc., Boca Raton, FL.

Noumonvi, K.D., Ferlan, M., Eler, K., Alberti, G., Peressotti, A. and Cerasoli, S., 2019. Estimation of carbon fluxes from eddy covariance data and satellite-derived vegetation indices in a karst grassland (Podgorski Kras, Slovenia). Remote sensing, 11(6), p.649.

Oleson, K., Lawrence, D.M., Bonan, G.B., Drewniak, B., Huang, M., Koven, C.D., Levis, S., Li, F., Riley, W., Subin, Z. and Swenson, S., 2013. Technical description of version 4.5 of the Community Land Model (CLM)(No. NCAR/TN-503+ STR). UCAR: Boulder, CO, USA.

Panek, J.A. and Goldstein, A.H., 2001. A controlled field experiment on the relation between drought stress, photosynthesis and stomatal conductance: implications for carbon storage and ozone deposition. Tree Physiology, 21, pp.337-344.

Palazzi, E., Mortarini, L., Terzago, S. and Von Hardenberg, J., 2019. Elevation-dependent warming in global climate model simulations at high spatial resolution. Climate Dynamics, 52(5), pp.2685-2702.

Papale, D., Black, T.A., Carvalhais, N., Cescatti, A., Chen, J., Jung, M., Kiely, G., Lasslop, G., Mahecha, M.D., Margolis, H. and Merbold, L., 2015. Effect of spatial sampling from European flux towers for estimating carbon and water fluxes with artificial neural networks. Journal of Geophysical Research: Biogeosciences, 120(10), pp.1941-1957.

Parton, W.J., Scurlock, J.M.O., Ojima, D.S., Gilmanov, T.G., Scholes, R.J., Schimel, D.S., Kirchner, T., Menaut, J.C., Seastedt, T., Garcia Moya, E. and Kamnalrut, A., 1993. Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide. Global biogeochemical cycles, 7(4), pp.785-809.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M. and Isaac, P., 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Scientific data, 7(1), pp.1-27.

Pei, Y., Dong, J., Zhang, Y., Yuan, W., Doughty, R., Yang, J., Zhou, D., Zhang, L. and Xiao, X., 2022. Evolution of light use efficiency models: Improvement, uncertainties, and implications. Agricultural and Forest Meteorology, 317, p.108905.

Pepin, N., Bradley, R.S., Diaz, H.F., Baraer, M., Caceres, E.B., Forsythe, N., Fowler, H., Greenwood, G., Hashmi, M.Z., Liu, X.D. and Miller, J.R., 2015. Elevation-dependent warming in mountain regions of the world. Nat Clim Change 5: 424–430.

Piao, S., Ciais, P., Friedlingstein, P., de Noblet-Ducoudré, N., Cadule, P., Viovy, N. and Wang, T., 2009. Spatiotemporal patterns of terrestrial carbon cycle during the 20th century. Global Biogeochemical Cycles, 23(4).

Piao, S., Sitch, S., Ciais, P., Friedlingstein, P., Peylin, P., Wang, X., Ahlström, A., Anav, A., Canadell, J.G., Cong, N. and Huntingford, C., 2013. Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO2 trends. Global change biology, 19(7), pp.2117-2132.

Potter, C.S., Randerson, J.T., Field, C.B., Matson, P.A., Vitousek, P.M., Mooney, H.A. and Klooster, S.A., 1993. Terrestrial ecosystem production: a process model based on global satellite and surface data. Global Biogeochemical Cycles, 7(4), pp.811-841.

Prentice, I.C., Liang, X., Medlyn, B.E. and Wang, Y.P., 2014. Reliable, robust and realistic: The three R's of next-generation land surface modelling. Atmospheric Chemistry and Physics Discussions, 14(17), pp.24811-24861.

Rangwala, I. and Miller, J.R., 2012. Climate change in mountains: a review of elevation-dependent warming and its possible causes. Climatic change, 114(3), pp.527-547.

Rawson, H.M., Begg, J.E. and Woodward, R.G., 1977. The effect of atmospheric humidity on photosynthesis, transpiration and water use efficiency of leaves of several plant species. Planta, 134(1), pp.5-10.

Real-Rangel, R.A., Pedrozo-Acuña, A. and Breña-Naranjo, J.A., 2018, December. Performance of precipitation and temperature datasets derived from GLDAS-2 and MERRA-2 in Mexico. In AGU Fall Meeting Abstracts.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A. and Grünwald, T., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. Global change biology, 11(9), pp.1424-1439.

Reichstein, M., Bahn, M., Ciais, P., Frank, D., Mahecha, M.D., Seneviratne, S.I., Zscheischler, J., Beer, C., Buchmann, N., Frank, D.C. and Papale, D., 2013. Climate extremes and the carbon cycle. Nature, 500(7462), pp.287-295.

Rodell, M., Houser, P.R., Jambor, U.E.A., Gottschalck, J., Mitchell, K., Meng, C.J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M. and Entin, J.K., 2004. The global land data assimilation system. Bulletin of the American Meteorological Society, 85(3), pp.381-394.

Rogers, A., 2014. The use and misuse of V c, max in Earth System Models. Photosynthesis research, 119(1), pp.15-29.

Rogers, A., Medlyn, B.E., Dukes, J.S., Bonan, G., Von Caemmerer, S., Dietze, M.C., Kattge, J., Leakey, A.D., Mercado, L.M., Niinemets, Ü. and Prentice, I.C., 2017. A roadmap for improving the representation of photosynthesis in Earth system models. New Phytologist, 213(1), pp.22-42.

Rui, H. and Beaudoing, H., 2011. Readme document for global land data assimilation system version 2 (GLDAS-2) products. GES DISC, 2011.

Running, S.W. and Hunt Jr, E.R., 1993. Generalization of a forest ecosystem process model for other biomes, BIOME-BCG, and an application for global-scale models.

Running, S.W., Thornton, P.E., Nemani, R. and Glassy, J.M., 2000. Global terrestrial gross and net primary productivity from the earth observing system. In Methods in ecosystem science (pp. 44-57). Springer, New York, NY.

Running, S.W., Nemani, R.R., Heinsch, F.A., Zhao, M., Reeves, M. and Hashimoto, H., 2004. A continuous satellite-derived measure of global terrestrial primary production. Bioscience, 54(6), pp.547-560.

Ryu, Y., Baldocchi, D.D., Kobayashi, H., van Ingen, C., Li, J., Black, T.A., Beringer, J., Van Gorsel, E., Knohl, A., Law, B.E. and Roupsard, O., 2011. Integration of MODIS land and atmosphere products with a coupled-process model to estimate gross primary productivity and evapotranspiration from 1 km to global scales. Global Biogeochemical Cycles, 25(4).

Saitoh, T.M., Nagai, S., Noda, H.M., Muraoka, H. and Nasahara, K.N., 2012. Examination of the extinction coefficient in the Beer–Lambert law for an accurate estimation of the forest canopy leaf area index. Forest Science and Technology, 8(2), pp.67-76.

Sellers, P.J., 1985. Canopy reflectance, photosynthesis and transpiration. International journal of remote sensing, 6(8), pp.1335-1372.

Sellers, P.J., Randall, D.A., Collatz, G.J., Berry, J.A., Field, C.B., Dazlich, D.A., Zhang, C., Collelo, G.D. and Bounoua, L., 1996. A revised land surface parameterization (SiB2) for atmospheric GCMs. Part I: Model formulation. Journal of climate, 9(4), pp.676-705.

Shuttleworth, W.J., 1993. Evaporation. In Handbook of Hydrology, Maidment, D.R., Ed., McGRaw Hill: New York, NY, USA, pp. 4.1–4.53.

Sims, D.A., Rahman, A.F., Cordova, V.D., El-Masri, B.Z., Baldocchi, D.D., Bolstad, P.V., Flanagan, L.B., Goldstein, A.H., Hollinger, D.Y., Misson, L. and Monson, R.K., 2008. A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS. Remote sensing of Environment, 112(4), pp.1633-1646.

Sitch, S., Smith, B., Prentice, I.C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J.O., Levis, S., Lucht, W., Sykes, M.T. and Thonicke, K., 2003. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. Global change biology, 9(2), pp.161-185.

Sitch, S., Huntingford, C., Gedney, N., Levy, P.E., Lomas, M., Piao, S.L., Betts, R., Ciais, P., Cox, P., Friedlingstein, P. and Jones, C.D., 2008. Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five Dynamic Global Vegetation Models (DGVMs). Global Change Biology, 14(9), pp.2015-2039.

Smith, T.M., Cramer, W.P., Dixon, R.K., Leemans, R., Neilson, R.P. and Solomon, A.M., 1993. The global terrestrial carbon cycle. Water, Air, and Soil Pollution, 70(1-4), pp.19-37.

Song, C. and Band, L.E., 2004. MVP: a model to simulate the spatial patterns of photosynthetically active radiation under discrete forest canopies. Canadian journal of forest research, 34(6), pp.1192-1203.

Song, C., Dannenberg, M.P. and Hwang, T., 2013. Optical remote sensing of terrestrial ecosystem primary productivity. Progress in Physical Geography, 37(6), pp.834-854.

Spitters, C.J.T., 1986. Separating the diffuse and direct component of global radiation and its implications for modeling canopy photosynthesis Part II. Calculation of canopy photosynthesis. Agricultural and Forest meteorology, 38(1-3), pp.231-242.

Still, C.J., Berry, J.A., Collatz, G.J. and DeFries, R.S., 2003. Global distribution of C3 and C4 vegetation: carbon cycle implications. Global biogeochemical cycles, 17(1), pp.6-1.

Subba, T., Gogoi, M.M., Pathak, B., Bhuyan, P.K. and Babu, S.S., 2020. Recent trend in the global distribution of aerosol direct radiative forcing from satellite measurements. Atmospheric Science Letters, 21(11), p.e975.

Sun, Z., Wang, X., Yamamoto, H., Tani, H., Zhong, G., Yin, S. and Guo, E., 2018. Spatial pattern of GPP variations in terrestrial ecosystems and its drivers: Climatic factors, CO2 concentration and land-cover change, 1982–2015. Ecological Informatics, 46, pp.156-165.

Sun, Z., Wang, X., Zhang, X., Tani, H., Guo, E., Yin, S. and Zhang, T., 2019. Evaluating and comparing remote sensing terrestrial GPP models for their response to climate variability and CO2 trends. Science of the total environment.

Testa, S., Soudani, K., Boschetti, L. and Mondino, E.B., 2018. MODIS-derived EVI, NDVI and WDRVI time series to estimate phenological metrics in French deciduous forests. International journal of applied earth observation and geoinformation, 64, pp.132-144.

Thornton, P.E., 2010. Biome-BGC version 4.2: Theoretical framework of Biome-BGC. Technical documentation.

Walker, A.P., Quaife, T., Van Bodegom, P.M., De Kauwe, M.G., Keenan, T.F., Joiner, J., Lomas, M.R., MacBean, N., Xu, C., Yang, X. and Woodward, F.I., 2017. The impact of alternative trait-scaling hypotheses for the maximum photosynthetic carboxylation rate (Vcmax) on global gross primary production. New Phytologist, 215(4), pp.1370-1386.

Wang, Y.P., 2003. A comparison of three different canopy radiation models commonly used in plant modelling. Functional Plant Biology, 30(2), pp.143-152.

Wang, Y.P. and Leuning, R., 1998. A two-leaf model for canopy conductance, photosynthesis and partitioning of available energy I:: Model description and comparison with a multi-layered model. Agricultural and Forest Meteorology, 91(1-2), pp.89-111

Wang, F., Wang, L., Koike, T., Zhou, H., Yang, K., Wang, A. and Li, W., 2011. Evaluation and application of a fine-resolution global data set in a semiarid mesoscale river basin with a distributed biosphere hydrological model. Journal of Geophysical Research: Atmospheres, 116(D21).

Wu, C., Munger, J.W., Niu, Z. and Kuang, D., 2010. Comparison of multiple models for estimating gross primary production using MODIS and eddy covariance data in Harvard Forest. Remote Sensing of Environment, 114(12), pp.2925-2939.

Xiao, Z., Liang, S., Wang, J., Song, J. and Wu, X., 2009. A temporally integrated inversion method for estimating leaf area index from MODIS data. IEEE Transactions on Geoscience and Remote Sensing, 47(8), pp.2536-2545.

Xiao, J., Zhuang, Q., Law, B.E., Chen, J., Baldocchi, D.D., Cook, D.R., Oren, R., Richardson, A.D., Wharton, S., Ma, S. and Martin, T.A., 2010. A continuous measure of gross primary production for the conterminous United States derived from MODIS and AmeriFlux data. Remote sensing of environment, 114(3), pp.576-591.

Xiao, J., Sun, G., Chen, J., Chen, H., Chen, S., Dong, G., Gao, S., Guo, H., Guo, J., Han, S. and Kato, T., 2013. Carbon fluxes, evapotranspiration, and water use efficiency of terrestrial ecosystems in China. Agricultural and forest meteorology, 182, pp.76-90.

Xie, X., Li, A., Jin, H., Tan, J., Wang, C., Lei, G., Zhang, Z., Bian, J. and Nan, X., 2019.

Assessment of five satellite-derived LAI datasets for GPP estimations through ecosystem models. Science of The Total Environment, 690, pp.1120-1130.

Yan, H., Wang, S.Q., Yu, K.L., Wang, B., Yu, Q., Bohrer, G., Billesbach, D., Bracho, R., Rahman, F. and Shugart, H.H., 2017. A Novel Diffuse Fraction-Based Two-Leaf Light Use Efficiency Model: An Application Quantifying Photosynthetic Seasonality across 20 AmeriFlux Flux Tower Sites. Journal of Advances in Modeling Earth Systems, 9(6), pp.2317-2332.

Yang, F., Lu, H., Yang, K., He, J., Wang, W., Wright, J.S., Li, C., Han, M. and Li, Y., 2017. Evaluation of multiple forcing data sets for precipitation and shortwave radiation over major land areas of China. Hydrology & Earth System Sciences, 21(11).

Yoder, R.E., Odhiambo, L.O. and Wright, W.C., 2005. Effects of vapor-pressure deficit and net-irradiance calculation methods on accuracy of standardized Penman-Monteith equation in a humid climate. Journal of irrigation and drainage engineering, 131(3), pp.228-237.

Yuan, H., Dai, Y., Xiao, Z., Ji, D. and Shangguan, W., 2011. Reprocessing the MODIS Leaf Area Index products for land surface and climate modelling. Remote Sensing of Environment, 115(5), pp.1171-1187.

Yuan, W., Liu, S., Zhou, G., Zhou, G., Tieszen, L.L., Baldocchi, D., Bernhofer, C., Gholz, H., Goldstein, A.H., Goulden, M.L. and Hollinger, D.Y., 2007. Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes. Agricultural and Forest Meteorology, 143(3-4), pp.189-207.

Yuan, W., Cai, W., Xia, J., Chen, J., Liu, S., Dong, W., Merbold, L., Law, B., Arain, A., Beringer, J. and Bernhofer, C., 2014. Global comparison of light use efficiency models for simulating terrestrial vegetation gross primary production based on the LaThuile database. Agricultural and Forest Meteorology, 192, pp.108-120.

Yuan, H., Dickinson, R.E., Dai, Y., Shaikh, M.J., Zhou, L., Shangguan, W. and Ji, D., 2014b. A 3D canopy radiative transfer model for global climate modeling: Description, validation, and application. Journal of climate, 27(3), pp.1168-1192.

Yuan, H., Dai, Y., Dickinson, R.E., Pinty, B., Shangguan, W., Zhang, S., Wang, L. and Zhu, S., 2017. Reexamination and further development of two-stream canopy radiative transfer models for global land modeling. Journal of Advances in Modeling Earth Systems, 9(1), pp.113-129.

Yuan, M., Leirvik, T. and Wild, M., 2021. Global trends in downward surface solar radiation from spatial interpolated ground observations during 1961–2019. Journal of Climate, 34(23), pp.9501-9521.

Zaehle, S., Sitch, S., Smith, B. and Hatterman, F., 2005. Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics. Global Biogeochemical Cycles, 19(3).

Zhang, F., Chen, J.M., Chen, J., Gough, C.M., Martin, T.A. and Dragoni, D., 2012. Evaluating spatial and temporal patterns of MODIS GPP over the conterminous US against flux measurements and a process model. Remote Sensing of Environment, 124, pp.717-729.

Zhang, W.L., Chen, S.P., Chen, J., Wei, L., Han, X.G. and Lin, G.H., 2007. Biophysical regulations of carbon fluxes of a steppe and a cultivated cropland in semiarid Inner Mongolia. Agricultural and Forest Meteorology, 146(3-4), pp.216-229.

Zhang, Y., Yu, G., Yang, J., Wimberly, M.C., Zhang, X., Tao, J., Jiang, Y. and Zhu, J., 2014. Climate-driven global changes in carbon use efficiency. Global Ecology and Biogeography, 23(2), pp.144-155.

Zhang, Y., Xiao, X., Wu, X., Zhou, S., Zhang, G., Qin, Y. and Dong, J., 2017. A global moderate resolution dataset of gross primary production of vegetation for 2000–2016. Scientific data, 4(1), pp.1-13.

Zhang Z, Chen JM, Guanter L, He L, Zhang Y. From canopy ‐ leaving to total canopy far ‐ red fluorescence emission for remote sensing of photosynthesis: First results from TROPOMI. Geophysical Research Letters. 2019 Nov 16;46(21):12030-40.

Zhao, M. and Running, S.W., 2010. Drought-induced reduction in global terrestrial net primary production from 2000 through 2009. science, 329(5994), pp.940-943.

Zhao, W. and Qualls, R.J., 2005. A multiple ‐ layer canopy scattering model to simulate shortwave radiation distribution within a homogeneous plant canopy. Water resources research, 41(8).

Zheng, Y., Shen, R., Wang, Y., Li, X., Liu, S., Liang, S., Chen, J.M., Ju, W., Zhang, L. and Yuan, W., 2020. Improved estimate of global gross primary production for reproducing its long-term variation, 1982–2017. Earth System Science Data, 12(4), pp.2725-2746.