https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

Open Access Earth System Science Data Discussions

Spatiotemporally consistent global dataset of the GIMMS Leaf Area Index (GIMMS LAI4g) from 1982 to 2020

Sen Cao[1,2,†] , Muyi Li[1,†], Zaichun Zhu[1,2,*], Junjun Zha[1], Weiqing Zhao[1], Zeyu Duanmu[1], Jiana Chen[1], Yaoyao Zheng[1], Yue Chen[1]

[1]School of Urban Planning and Design, Shenzhen Graduate School, Peking University, Shenzhen 518055, China
[2]Key Laboratory of Earth Surface System and Human—Earth Relations, Ministry of Natural Resources of China, Shenzhen Graduate School, Peking University, Shenzhen 518055, China.
[†]These authors contributed equally to this work

Correspondence: Zaichun Zhu (zhu.zaichun@pku.edu.cn)

Abstract: Leaf Area Index (LAI) with an explicit biophysical meaning is a critical variable to characterize terrestrial ecosystems. Long-term global datasets of LAI have served as fundamental data support for monitoring vegetation dynamics and exploring its interactions with other Earth components. However, current LAI products face several limitations associated with spatiotemporal consistency. In this study, we employed the Back Propagation Neural Network (BPNN) and a data consolidation method to generate a new version of the half-month 1/12° Global Inventory Modeling and Mapping Studies (GIMMS) LAI product, i.e., GIMMS LAI4g, for the period 1982−2020. The significance of the GIMMS LAI4g was the use of the latest PKU GIMMS NDVI product and 3.6 million high-quality global Landsat LAI samples to remove the effects of satellite orbital drift and sensor degradation and to develop spatiotemporally consistent BPNN models. The results showed that the GIMMS LAI4g exhibited higher accuracy than its predecessor (GIMMS LAI3g) and two mainstream LAI products (Global LAnd Surface Satellite [GLASS] LAI and Long-term Global Mapping [GLOBMAP] LAI), with an R^2 of 0.95, mean absolute error of 0.18 m^2/m^2, and mean absolute percentage error of 15% which meet the accuracy target proposed by the Global Climate Observation System. It outperformed other LAI products for most vegetation biomes in a majority area of the land. It efficiently eliminated the effects of satellite orbital drift and sensor degradation and presented a better temporal consistency before and after the year 2000 and a more reasonable global vegetation trend. The GIMMS LAI4g product could potentially facilitate mitigating the disagreements between studies of the long-term global vegetation changes and could also benefit the model development in Earth and environmental sciences.

Keywords: GIMMS LAI4g; PKU GIMMS NDVI; Landsat LAI samples; MODIS LAI; BPNN

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

Open Access Earth System Science Data Discussions

1. Introduction

Leaf Area Index (LAI), defined as half of the total green leaf area per unit of horizontal surface, is a key variable in vegetation change monitoring (Piao et al., 2015; Valderrama-Landeros et al., 2016; Zhu et al., 2016), land surface modeling (Boussetta et al., 2013; Boussetta et al., 2015; Chen et al., 2015), crop yield estimation (De Wit et al., 2012; Dente et al., 2008), etc. It is one of the basic terrestrial climate variables selected by the Global Climate Observation System (GCOS) (Gcos, 2011). Remote sensing observation has been the only reliable means of obtaining spatiotemporally continuous LAI products at the global scale (Ma and Liang, 2022). The common practice is to relate remote sensing data with ground LAI measurements or other remote sensing products of higher reliability (as LAI reference), using methods including statistical modeling (Liu et al., 2012; Kimura et al., 2004; Broge and Leblanc, 2001), physical modeling (Myneni et al., 2002), and machine learning (Xiao et al., 2014; Zhu et al., 2013; Ma and Liang, 2022; Kang et al., 2021). Over past years, a number of long-term global LAI products, such as the third generation Global Inventory Modeling and Mapping Studies (GIMMS) LAI (GIMMS LAI3g) (Zhu et al., 2013), the Global LAnd Surface Satellite (GLASS) LAI (Xiao et al., 2016), the Long-term Global Mapping (GLOBMAP) LAI (Liu et al., 2012), and the Terrestrial Climate Data Record (TCDR) LAI (Claverie et al., 2016), have been released. These products have provided many in-depth insights into how global vegetation responds to human disturbances and global warming (Zhu et al., 2016; Piao et al., 2015; Chen et al., 2019a). Specifically, the GIMMS LAI3g has been one of core data reference in IPCC Sixth Assessment Report for the assessment of global vegetation changes (Eyring et al., 2021).

However, the accuracies of the current LAI products have been limited by uncertainties primarily in the remote sensing data and the LAI reference data (Fang et al., 2019). First, remote sensing data being used have some common issues. For example, false gradual signals and mutations have been widely observed in the LAI time series prior to the late 1990s for mainstream long-term LAI products, as most of them utilized data from the Advanced Very High Resolution Radiometer (AVHRR) (Wang et al., 2022). The AVHRR sensors onboard National Oceanic and Atmospheric Administration (NOAA) satellites were the only remote sensing data sources before the late 1990s that provided spatiotemporally continuous observations over the globe. Nevertheless, they suffered from issues of NOAA satellite orbital drift and AVHRR sensor degradation, particularly in the tropical area of evergreen broadleaf forests (Pinzon and Tucker, 2014). Second, the LAI reference data used to build LAI models have been scarce, particularly before the late 1990s. After the year of 2000, global LAI products became increasingly available from advanced sensors such as the Moderate Resolution Imaging Spectroradiometer (MODIS) (Myneni et al., 2002), the Système Pour l'Observation de la Terre (SPOT) (Baret et al., 2007), and the Visible Infrared Imaging Radiometer Suite (VIIRS) (Yan et al., 2018). These LAI products have been elaborately and collaboratively validated despite with a short time span. Current studies employed them as LAI reference to build post-2000 AVHRR−LAI models and extrapolated the models onto pre-2000 AVHRR data so that long-term LAI product could be produced (Chen et al., 2019a). Nonetheless, the legality of the extrapolation remains questioned since the AVHRR−LAI relationship could change with time.

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

The uncertainties in the remote sensing and LAI reference data, together with the differences in modeling algorithms, have led the performance of long-term global LAI products to vary from one to another. Inconsistencies were continually found between LAI products regardless of the remote sensing data source used (Jiang et al., 2017). For example, four popular global data sets of LAI (1982−2010s), namely the GLASS LAI, GLOBMAP LAI, GIMMS LAI3g, and TCDR LAI, showed significant differences in LAI trends, interannual variabilities, and uncertainty variations (Jiang et al., 2017; Xiao et al., 2017). In tropical areas, the average LAI difference can be up to one unit (Yan et al., 2016). These differences between LAI products have raised many concerns about the robustness of existing vegetation change analysis and land surface modeling (Alkama et al., 2022; Jiang et al., 2017; Piao et al., 2015).

Recent advances in land data products have provided pathways to address the uncertainties. In particular, the PKU GIMMS Normalized Difference Vegetation Index (NDVI) product (1982−2020) by Li et al. (in review) efficiently eliminated the evident NOAA orbital drift and AVHRR sensor degradation effects. It demonstrates higher accuracy than the predecessor (GIMMS NDVI3g) and shows a high temporal consistency with MODIS NDVI. Zha et al. (in preparation) compiled a set of global reference LAI (before the year 2000) and created over 6 million high-quality Landsat LAI samples over the globe from 1984 to 2020 (also see Zha, 2022). The validation against LAI field measurements showed an R^2 of 0.76. Although these Landsat LAI samples can hardly be used to characterize the global vegetation change because they are not spatiotemporally continuous, they can serve as reliable LAI references.

In this context, the objective of this study is to derive a new generation of GIMMS LAI products (GIMMS LAI4g, 1982−2020) using machine learning models based on the PKU GIMMS NDVI product and massive high-quality Landsat LAI samples and a data consolidation method based on the Reprocessed MODIS LAI product. We employ the PKU GIMMS NDVI and the Landsat LAI samples to address the uncertainties in remote sensing and LAI reference data. With these data, biome-specific Back Propagation Neural Network (BPNN) models are developed with additional explanatory variables (the longitude and latitude, the NDVI month, and the NOAA number and years since launch). The GIMMS LAI4g product is then generated from the BPNN models. Finally, the GIMMS LAI4g is consolidated with the Reprocessed MODIS NDVI product via a pixel-wise linear fusion method to extend the temporal coverage to the year 2020. We evaluate the GIMMS LAI4g's accuracy by a direct validation method and compare its accuracy to those of three other global LAI products, i.e., GIMMS LAI3g, GLASS LAI, and GLOBMAP LAI. The temporal consistency of the global LAI products and their LAI trends are also analyzed.

2. Data

A total of seven global datasets were used in this study, namely, the PKU GIMMS NDVI, Landsat LAI sample dataset, MODIS Land-Cover Type, Reprocessed MODIS LAI, GLASS LAI, GLOBMAP LAI, and GIMMS LAI3g. The PKU GIMMS NDVI was the primary data source from which the GIMMS LAI4g was generated. The Landsat LAI sample dataset was used

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

as the LAI reference in machine learning model establishment and product evaluation. The MODIS Land-Cover Type product provided vegetation biome types in the LAI modeling. The Reprocessed MODIS LAI was used to extend the temporal coverage of the GIMMS LAI4g. The GLASS LAI, GLOBMAP LAI, and GIMMS LAI3g are three mainstream global LAI products that were included for an inter-comparison purpose.

2.1 PKU GIMMS NDVI

The PKU version of the GIMMS NDVI product (PKU GIMMS NDVI) was employed in this study (Li et al., in review). It has a spatial resolution of 1/12° and a temporal resolution of half-month. The PKU GIMMS NDVI was generated from the GIMMS NDVI3g by machine learning models that employed massive high-quality and global-wide Landsat NDVI samples and a data consolidation method that employed the MODIS NDVI product. The major improvement of PKU GIMMS NDVI over its counterparts is that it well removed the NOAA orbital drift and AVHRR sensor degradation effects. Its overall R^2, Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) is 0.97, 0.03 m^2/m^2, and 9%, respectively. PKU GIMMS NDVI inherited the quality control (QC) information from the GIMMS NDVI3g. A QC value of 0, 1, and 2 indicates NDVI of good quality, NDVI retrieved from spline interpolation, and NDVI retrieved from average seasonal profile, respectively. The PKU GIMMS NDVI record during AVHRR missions from 1982 to 2015 was used in this study. It is available at https://zenodo.org/record/7441559#.Y7J7y3ZByCo.

2.2 Landsat LAI sample dataset

The Landsat LAI sample dataset provides approximately 6 million high-quality samples with a spatial resolution of 1/12 ° and a temporal resolution of half a month. It covers the global vegetated area with all vegetation biome types defined in the MODIS land cover product (the third classification scheme; see section 2.4) and a long-time span from 1984 to 2020. The samples were derived by relating carefully refined Landsat 5/7/8 reflectance to MODIS LAI via the Random Forests algorithm (Zha, in preparation). Its validation at the BEnchmark Land Multisite ANalysis and Intercomparison of Products (BELMANIP) sites and inter-comparison with MODIS LAI showed a high accuracy of estimation (R^2 = 0.76, MAE = 0.45 m^2/m^2, Root Mean Square Error (RMSE) = 0.66 m^2/m^2) and a high temporal consistency. This study selected 3.6 million of Landsat LAI samples between 1984 and 2015.

2.3 Reprocessed MODIS LAI product

The latest version of the Reprocessed MODIS LAI product (version 6) has a time span of 2000−2020, a temporal resolution of 8-day or one-month, and a spatial resolution ranging from 500 m to 0.5°. The product was derived from the MODIS LAI Version 6 products (Myneni, 2015b, a) and MODIS Land Cover Type product (Friedl, 2019) using an integrated two-step method (Yuan et al., 2011). Compared to the original MODIS LAI products, it is more spatiotemporally continuous

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

and consistent as verified by 44 LAI reference maps which contain true LAI values collected over a subset of 26 ground sites. We downloaded the 8-day 0.05° data from http://globalchange.bnu.edu.cn/research/laiv6#download, and resampled the data to have the same spatial resolution (1/12°) and temporal resolution (half a month) as the GIMMS LAI4g. The temporal subset of 2004−2020 was used in this study because the LAI data in the evergreen broadleaf forest were found exceptional low between 2000 and 2003 than other years (Figure A1).

2.4 MODIS Land-Cover Type product (MCD12Q1)

The MODIS Land Cover Type Product (MCD12Q1, version 6) supplies global maps of annual land cover with a spatial resolution of 500 m between 2001 and 2019 (Friedl et al., 2002; Friedl, 2019). It includes five legacy classification schemes. This study selected the third classification scheme (Annual LAI classification). The Annual LAI classification scheme includes eight natural vegetation types (Evergreen Needleleaf Forests [ENF], Evergreen Broadleaf Forests [EBF], Deciduous Needleleaf Forests [DNF], Deciduous Broadleaf Forests [DBF], SHRublands [SHR], SAVannas [SAV], GRAsslands [GRA], CROplands [CRO]), and three non-vegetated lands (WATer bodies [WAT], Non-VeGetated lands [NVG], and URban and Built-up lands [URB]). This study also used [GLO] in data analysis to represent the global vegetation biome (the ensemble of the eight vegetation types). The spatial resolution of MCD12Q1 was spatially aggregated to 1/12° in this study to match that of PKU GIMMS NDVI. For each 1/12° grid, the aggregation was conducted by calculating frequencies of each biome type between 2001 and 2019 and identifying the most frequent one. This generated a global land cover map that was considered static from 1982 through 2020 in this study (Figure B1). With potential errors, this strategy could be the best option at the time.

2.5 GLASS LAI

The GLASS LAI (version 5) with a temporal resolution of 8 days was generated from the 0.05° resolution NOAA/AVHRR surface reflectance dataset provided by NASA's Long Term Data Record (LTDR) project (1982−2000) and the 1 km resolution Terra/MODIS surface reflectance dataset (MOD09) (2000−2018) (Xiao et al., 2016). In the algorithm, biome-specific general regression neural networks were built between the surface reflectance data and LAI reference data which were created by fusing Terra/MODIS LAI (MOD15) with clump-corrected CYCLOPES LAI over BELMANIP sites (Xiao et al., 2016). The neural networks were then used to predict global LAI (Xiao et al., 2014). The GLASS LAI (V5) product was acquired from ftp://ftp.glcf.umd.edu/. It should be noted that version 6 of the GLASS LAI product (250-m resolution) is currently available, but only for years after 2000 (Ma and Liang, 2022).

2.6 GLOBMAP LAI

The latest GLOBMAP LAI product (version 3) with a spatial resolution of 1/13.75° and temporal resolutions of half-month (1982−2000) or 8-day (2001−now) was generated based on GIMMS NDVI product (1982−2000) (Tucker et al., 2005)

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

and Terra/MODIS surface reflectance (MOD09A1 C6) (2001−now). The algorithm established relationships between MODIS LAI and GIMMS NDVI in a pixel-wised manner during their overlapping period of 2000−2006. The relationships were then applied to GIMMS NDVI before 2000 (Liu et al., 2012). The GLOBMAP LAI (V3) product was acquired from https://zenodo.org/ record/4700264#.YH6QD-gzbOk.

2.7 GIMMS LAI3g

The GIMMS LAI3g product (version 4) (1982−2016) was generated biweekly in a 1/12° spatial resolution (Zhu et al., 2013). It is available at https://drive.google.com/drive/folders/0BwL88nwumpqYaFJmR2poS0d1ZDQ?resourcekey=0-9IRE9s-0tFGfwB5qTpLjZw&usp=sharing/. The algorithm related the GIMMS' third-generation NDVI (NDVI3g) to MODIS LAI of Beijing Normal University (BNU) version via feed-forward neural networks (Yuan et al., 2011). Twelve neural networks, one for each month, were built using monthly averaged LAI and NDVI data between 2000 and 2009. The GIMMS LAI3g was then produced from GIMMS NDVI3g by applying the neural networks to the period of 1982 to 2016.

3. Methodology

The methodology includes three key steps:1) generating the GIMMS LAI4g product from biome-specific Back Propagation Neural Network (BPNN) models based on PKU GIMMS NDVI, Landsat LAI sample, and other explanatory variables; 2) consolidating the GIMMS LAI4g product with the MODIS LAI product using a pixel-wise linear fusion method in their overlapping timespan (2004−2015); and 3) evaluating the GIMMS LAI4g product using Landsat LAI samples and comparing it with other global LAI products.

3.1 Generation of GIMMS LAI4g using Back Propagation Neural Network (BPNN)

The artificial neural network (ANN) is a machine-learning algorithm inspired by the structure and function of biological neural networks (Basheer and Hajmeer, 2000; Zhang et al., 1998). It has been frequently used in ecological studies and the generation of global LAI products (Panda et al., 2010; Jahan and Gan, 2011; Zhu et al., 2013; Xiao et al., 2014; Claverie et al., 2016). For example, a typical ANN, general regression neural network, and Back Propagation Neural Network (BPNN) was employed in the production of TCDR LAI (version 5) (Claverie et al., 2016), GLASS LAI (version 5) (Xiao et al., 2014), and GIMMS LAI3g (Zhu et al., 2013), respectively. A typical ANN comprises input, output, and hidden layers, with each containing several artificial neurons. During the model training process, signals flowed from the input layer to the output layer, after likely passing through several hidden layers. Errors in the output layer propagate backward to the previous layers until they satisfy the user-defined threshold, and the network attempts to minimize the discrepancies between observations and predictions (Basheer and Hajmeer, 2000; Zhang et al., 1998).

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

This study used the BPNN model to predict LAI values from the PKU GIMMS NDVI (1982−2015). Individual BPNN models were developed for each vegetation biome. The target variable in BPNN models was mainly from the Landsat LAI samples (1984−2015) but also included a small number of MODIS LAI values in regions where Landsat LAI samples were lacking. These regions were mostly located in northern high latitudes that suffer from polar night phenomena and low solar altitude angles in the winter. Corresponding PKU GIMMS NDVI values of the same time and at the same locations with the LAI samples were extracted as the explanatory variable. The LAI samples and associated PKU GIMMS NDVI were further refined. Locations with negative NDVI values (e.g., contaminated by snow and inland water bodies) and non-zero QC values in the PKU GIMMS NDVI product were removed. After the refinement, the samples were randomly divided into two groups, i.e., the dataset for BPNN construction (80%) and the dataset for LAI product evaluation (20%).

In the BPNN models, we also incorporated spatial information (the longitude and latitude at the sample location), temporal information (the NDVI month), and NOAA satellite information (NOAA satellite number and years since launch; not applicable for MODIS LAI) as additional explanatory variables. A stepwise method was employed to determine the best combination of explanatory variables for each vegetation biome. The PKU GIMMS NDVI data were first included and evaluated in the BPNN models (Scenario 1 or S1). Then, the spatial information that accounts for spatial autocorrelation (S2), temporal information that accounts for vegetation dynamic (S3), NOAA satellite number (S4), and years since NOAA launch (S5) that accounts for potential satellite and sensor issues were added one by one. In model establishment, we repeatedly (50 times) selected 50,000 random samples with replacement for each vegetation biome. The 50,000 samples were split into 90% for model training and 10% for model evaluation, in which three error metrics of R^2, MAE, and MAPE, were calculated. The error metrics determined the optimum combination of explanatory variables and the optimum parameters for the final BNPP model of each biome.

3.2 Consolidation of GIMMS LAI4g and MODIS LAI

The GIMMS LAI4g product derived from the PKU GIMMS NDVI (1982−2015) which was based on AVHRR data did not include LAI data since 2015. As such, it can hardly be used to characterize recent vegetation dynamics. A couple of global products have provided up-to-date LAI data using satellite sensors available since the late 1990s (Baret et al., 2007). A common practice to generate the long-term LAI product is to consolidate the AVHRR-based LAI product with the post-2000 LAI product. For example, both GLASS LAI and GLOBMAP LAI consolidated LAI products from AVHRR and MODIS. MODIS has been one of the most popular and verified data sources for LAI production. In this study, the Reprocessed MODIS LAI product (2001−2020) was employed to extend the time span of the GIMMS LAI4g.

The consolidation adopted the pixel-wise linear fusion method proposed by Mao et al. (2012). In the method, LAI values in the overlapping period of 2004−2015 were extracted from the GIMMS LAI4g and the MODIS LAI. Then, the most appropriate linear regression equation was determined from the LAI values at each pixel location, with GIMMS LAI4g data as the explanatory variable and MODIS LAI data as the target variable. Compared to the global or biome-specific regression

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

models, this pixel-wise method has demonstrated excellent accuracies, especially in regulating the temporal consistency between datasets (Mao et al., 2012). The final LAI product comprised the GIMMS LAI4g (after consolidation) (1982−2003) and the Reprocessed MODIS LAI (2004−2020).

3.3 Evaluation of the GIMMS LAI4g product

In this study, the LAI reference samples were evaluated in terms of their number, spatial distribution, and temporal distribution under different vegetation biome types. To assess the representativeness of the samples, we also compared LAI reference values at the sample locations to those from the GIMMS LAI3g, GLASS LAI, and GLOBMAP LAI using a frequency histogram.

The performance of our GIMMS LAI4g product was evaluated and compared with three other global long-term LAI products (i.e., the GIMMS LAI3g, GLASS LAI, and GLOBMAP LAI) using Landsat LAI samples. The evaluation was conducted for each vegetation biome using the remaining 20% Landsat sample points (with a spatial resolution of 1/12°). Three measures of error were used: R^2, MAE (m^2/m^2), and MAPE (%). R^2 measures the percentage of variations that models can explain; and MAE and MAPE measure absolute and relative error values at the sample level. For inter-comparison between the GIMMS LAI4g and other LAI products, the spatial resolution and temporal resolution of all LAI products have been unified to 1/12° and half a month, respectively. We used a dominance map to demonstrate the global distribution of products with higher accuracy. The map was drawn with 2° × 2° grids using MAE values from the GIMMS LAI4g, GIMMS LAI3g, and GLASS LAI. The color of each grid was composed of reciprocal MAE averages of the LAI products, i.e., a lower MAE average can lead to a higher weight in the composite. The GLOBMAP LAI was not included because of its much higher MAE than other products. We also showcased the spatial consistencies between the four global LAI products by their spatial average along latitude (at an interval of 1°) in January and July of the years 1990, 2000, and 2010.

The temporal consistency of the GIMMS LAI4g was evaluated from three perspectives. First, LAI bias was used to examine whether the NOAA orbital drift and AVHRR sensor degradation effects were alleviated in the tropical area of EBF. The bias was calculated as the mean value of LAI deviation relative to Landsat LAI in percentage (Helder et al., 2013). If there is orbital drift or sensor degradation, the bias will drastically fluctuate; otherwise, it remains constant. Seasonal fluctuations in the time series of NDVI bias were first removed via the multi-year averaging method. Then, inter-annual trends of the bias were extracted via the Ensemble Empirical Mode Decomposition (EEMD) approach (Huang et al., 1998). Second, the efficiency of data consolidation between the GIMMS LAI4g and MODIS LAI was reported. We also checked whether the data consolidation altered the self-consistency of the GIMMS LAI4g product over time. Third, we used the Landsat LAI samples as the reference to evaluate the consistency of the GIMMS NDVI4g between different periods (p1: 1984−2015; p2: 1984−2000; and p3: 2001-2015) and compared the consistency with other three LAI products. The consistency was quantified by temporal changes in R^2, MAE, and MAPE.

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

The LAI trends between 1982 and 2015 were derived and compared between the four global LAI products. Linear regression analysis was performed on the LAI time series at the pixel level. The trend was calculated as the slope of the fitting line, which shows the vegetation trend of greening (positive slope) or browning (negative slope). This produced a global map of LAI trending. We also analyzed annual LAI variations and calculated annual LAI trends for all vegetation biome types. The annual LAI value is an area-weighted average based on the vegetation biome type.

4. Results

4.1 Examination of LAI reference data

The spatial and temporal distributions of the LAI reference data were determined mostly by the availability of Landsat images but also by the occurrence of cloud cover, aerosol, and other factors. Figure 1 shows the spatial distribution of the 3.6 million LAI samples primarily from the Landsat LAI dataset. The sample size for each vegetation biome, ranging from 116,873 (ENF) to 1,503,768 (GRA) is also listed. The sample locations spanned all latitudes of the vegetated area (Figure 1a), and no samples were selected from non-vegetation regions. The sample count per biome was approximately proportionate to the biome area (Figure 1b). In northern high latitudes, Landsat images were scarce throughout the winter due to the polar night phenomenon and the low solar altitude angle; and in the tropical area, Landsat images were frequently contaminated by precipitation and clouds. As a result, the number of available samples was limited in these two areas (Figure 1a). We addressed this issue by introducing 40,000 samples from the Reprocessed MODIS LAI at locations and months that Landsat LAI samples were scarce.

To evaluate the representativeness of the Landsat LAI samples, we calculated a frequency histogram based on all Landsat LAI sample values between 1984−2015 and compared it to those based on GIMMS LAI3g, GLASS LAI, and GLOBMAP LAI (Figure 1c). During 1984−2015, the LAI value distribution in the Landsat samples was similar to those in the other three products at global vegetation pixels (Fig. 1c), indicating that the Landsat LAI samples used in this study have good representativeness.

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

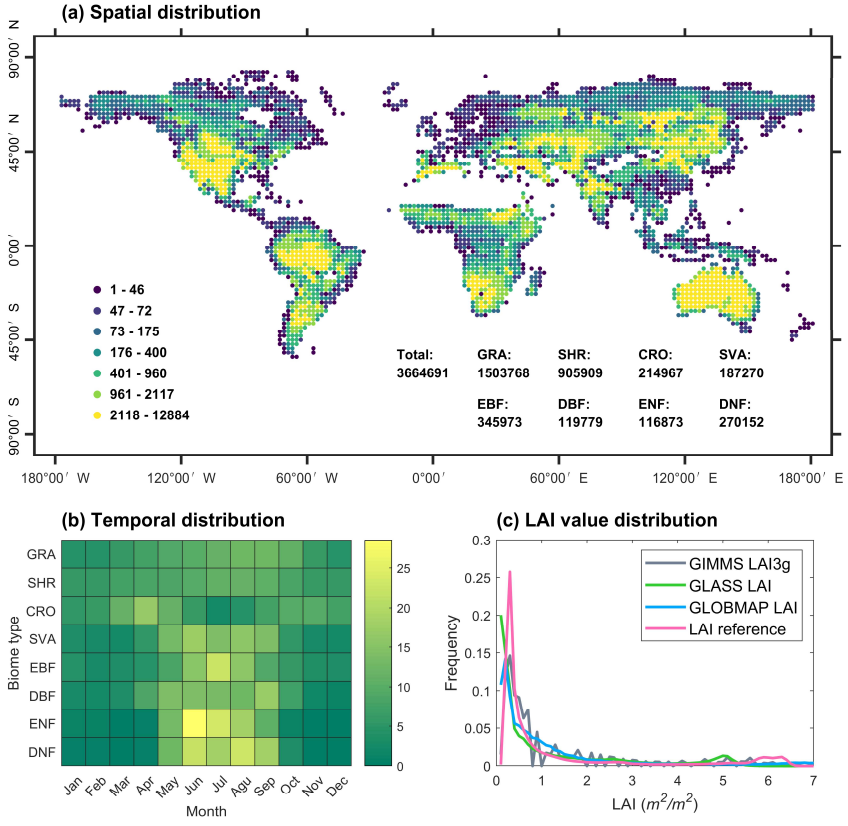


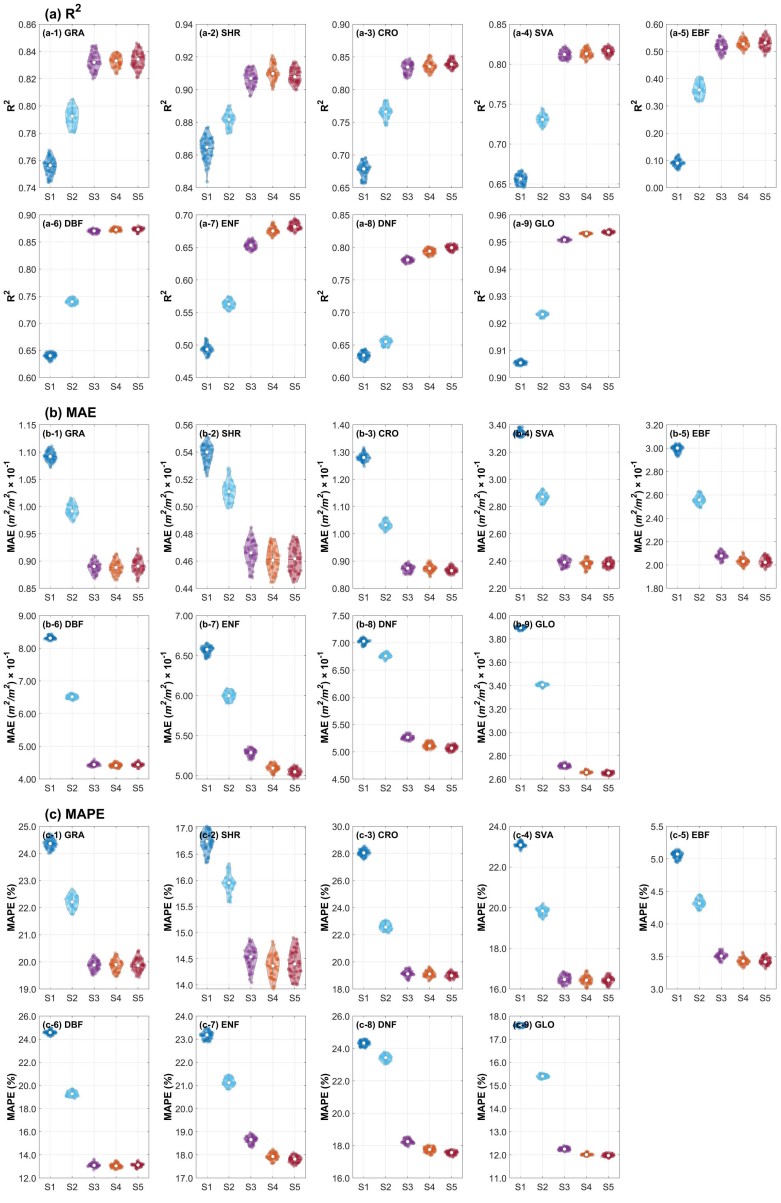
Figure 1. Spatial and temporal distribution of the LAI reference data. (a) The global distribution of LAI samples in 2° grids. The LAI sample size for each vegetation is listed. (b) The temporal distribution of LAI samples for the eight vegetation biome types. (c) The distribution of LAI values in percentage (bin width = 0.1) for Landsat LAI samples, GIMMS LAI3g, GLASS LAI, and GLOBMAP LAI. It should be noted that 40,000 Reprocessed MODIS LAI samples were introduced at locations and months that Landsat LAI samples were scarce.

4.2 The optimum BPNN models

For each vegetation biome, different combinations of explanatory variables (S1 to S5, see section 3.1) were tested in BNPP models. The variations in accuracy are shown in Figure 2 and Table 1. The inclusion of spatial information and temporal information has significantly improved the model performance with much higher R^2, lower MAE, and lower MAPE for most vegetation biomes (Figure 2). The improvement from spatial information was slight for DNF, probably because of its relative

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

concentration in certain middle- and high-latitudes of Eurasia (Figure B1). The inclusion of NOAA satellite number and years since its launch brought subtle but discernible improvements towards the accuracy of BPNN models.



https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

Figure 2. Performance of different combinations of explanatory variables (S1 to S5) in BPNN models for each vegetation biome. (a), (b) and (c) shows the R^2, MAE, and MAPE, respectively, calculated based on Landsat LAI samples. GLO represents the global vegetation biome. The combinations of explanatory variables are (S1) NDVI alone; (S2) NDVI and spatial information (longitude and latitude); (S3) NDVI, spatial information, and temporal information (month); (S4) NDVI, spatial information, temporal information, and NOAA satellite number; and (S5) NDVI, spatial information, temporal information, NOAA satellite number and years since its launch.

Using all explanatory variables in the BPNN model (S5) has resulted in R^2 values > 0.80 for most biome types except EBF (0.52) and ENF (0.72) and MAPE values < 20% (Table 1). MAPE of EBF was as low as 3.30%. For the global vegetation biome as a whole, accuracies of the BPNN model in S5 were R^2 of 0.96, MAE of 0.25 m^2/m^2, and MAPE of 11.35% (Table 1). As such, the BPNN models adopted the combination of all explanatory variables, including NDVI, longitude, latitude, month, NOAA satellite number, and NOAA satellite in orbit duration.

Table 1 Error metric values for different combinations of explanatory variables (S1 to S5) in BPNN of each vegetation biome. Values in this table correspond to Figure 2. The combinations of explanatory variables are (S1) NDVI alone; (S2) NDVI and spatial information (longitude and latitude); (S3) NDVI, spatial information, and temporal information (month); (S4) NDVI, spatial information, temporal information, and NOAA satellite number; and (S5) NDVI, spatial information, temporal information, NOAA satellite number and years since its launch. GLO represents the global vegetation biome.

Metrics	Combinations	Biome type								
		GRA	SHR	CRO	SVA	EBF	DBF	ENF	DNF	GLO
R^2	S1	0.76	0.86	0.68	0.66	0.09	0.64	0.49	0.63	0.91
	S2	0.79	0.88	0.77	0.73	0.36	0.74	0.56	0.65	0.92
	S3	0.83	0.91	0.83	0.81	0.52	0.87	0.65	0.78	0.95
	S4	0.83	0.91	0.84	0.81	0.53	0.87	0.68	0.79	0.95
	S5	0.83	0.91	0.84	0.82	0.53	0.87	0.68	0.80	0.95
MAE (m^2/m^2)	S1	0.11	0.05	0.13	0.33	0.30	0.83	0.66	0.70	0.39
	S2	0.10	0.05	0.10	0.29	0.26	0.65	0.60	0.68	0.34
	S3	0.09	0.05	0.09	0.24	0.21	0.45	0.53	0.53	0.27
	S4	0.09	0.05	0.09	0.24	0.20	0.44	0.51	0.51	0.27
	S5	0.09	0.05	0.09	0.24	0.20	0.45	0.50	0.51	0.27
MAPE (%)	S1	24.36	16.75	28.06	23.09	5.06	24.60	23.16	24.33	17.61
	S2	22.18	15.95	22.59	19.82	4.33	19.29	21.16	23.40	15.40
	S3	19.85	14.55	19.17	16.50	3.51	13.15	18.64	18.26	12.27
	S4	19.88	14.36	19.11	16.45	3.44	13.09	17.94	17.73	12.02
	S5	19.88	14.45	19.03	16.46	3.43	13.16	17.79	17.56	11.98

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

4.3 Validation of the GIMMS LAI4g and other LAI products

Figure 3 shows the validation results for the four global LAI products (our GIMMS LAI4g, GIMMS LAI3g, GLASS LAI, and GLOBMAP LAI) using the remaining 20% Landsat sample points. In general, GIMMS LAI4g (R^2 = 0.95, MAE = 0.18 m^2/m^2, MAPE = 15%) had the highest accuracy, followed by GIMMS LAI3g (R^2 = 0.92, MAE = 0.26 m^2/m^2, MAPE = 22%) and GLASS LAI (R^2 = 0.91, MAE = 0.29 m^2/m^2, MAPE = 24%). GLOBMAP LAI (R^2 = 0.77, MAE = 0.46 m^2/m^2, MAPE = 39%) had the lowest accuracy. The MAPE value of 15% in GIMMS LAI4g achieves the LAI accuracy target proposed by GCOS.

Regarding individual vegetation biome types, the GIMMS LAI4g product outperformed the other three except for DBF. In DBF, the GLASS LAI (R^2 = 0.75, MAE = 0.70 m^2/m^2, MAPE = 21%) has slightly higher accuracies than the GIMMS LAI4g (R^2 = 0.67, MAE = 0.74 m^2/m^2, MAPE = 22%). The most accurate vegetation biome varied with error metrics for all LAI products. In the GIMMS LAI4g, GIMMS LAI3g, and GLOBMAP LAI products, SHR had the highest accuracies in R^2 and MAE (R^2 = 0.90, 0.74, and 0.57, respectively; MAE = 0.06 m^2/m^2, 0.09 m^2/m^2, and 0.20 m^2/m^2, respectively); meanwhile, DBF had the highest accuracies in MAPE (MAPE = 4%, 10%, and 24%, respectively). The most accurate vegetation biome in GLASS LAI could be DBF, SHR, or EBF, determined by R^2, MAE, or MAPE, respectively. This discrepancy was attributed to the nature of the error metrics. For instance, EBF with higher absolute LAI values generally produced the lowest MAPE. However, the R^2 and MAE proposed that EBF could be the most inaccurate vegetation biome (R^2 = 0.59, 0.07, 0.20, and 0.05, respectively; MAE = 0.22 m^2/m^2, 0.61 m^2/m^2, 0.77 m^2/m^2, and 1.4 m^2/m^2, respectively). The LAI accuracy in EBF was low because it is primarily distributed in tropical areas where the quality of remote sensing data is poor owing to frequent clouds and rains.

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

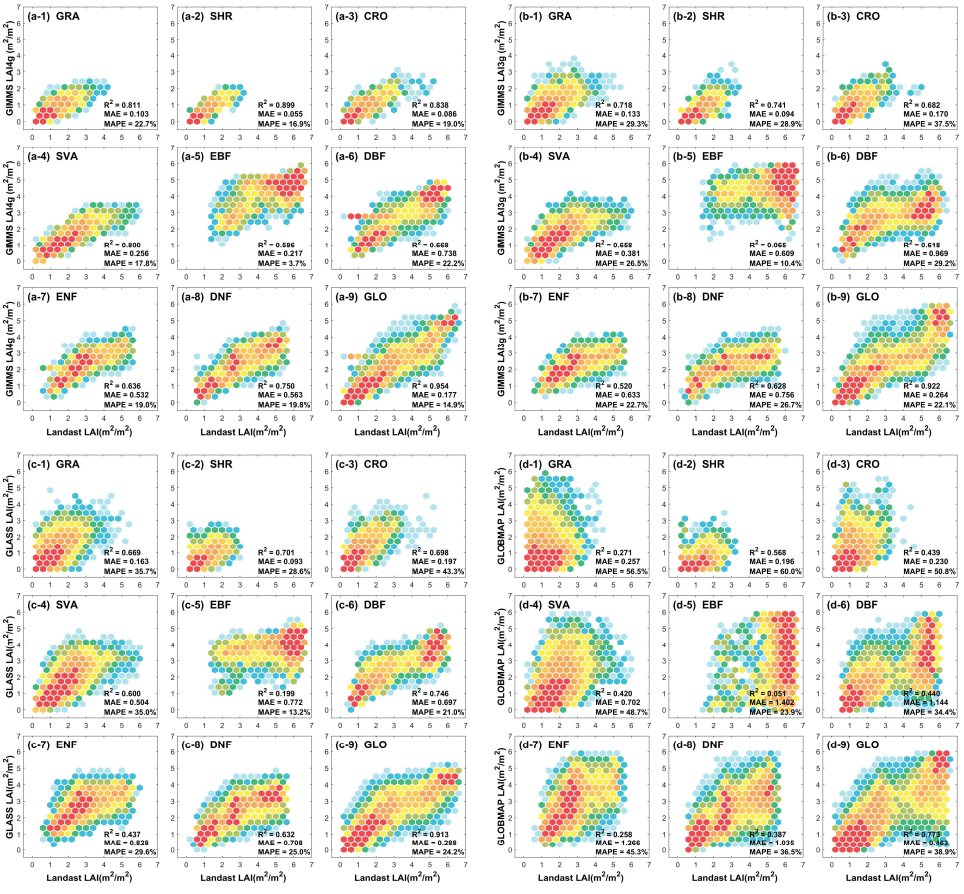


Figure 3. Validation of the (a) GIMMS LAI4g, (b) GIMMS LAI3g, (c) GLASS LAI, and (d) GLOBMAP LAI products in different vegetation biomes using Landsat LAI samples from 1984 to 2015. The error metrics are R^2, MAE, and MAPE. GLO represents the global vegetation biome. The color of the dots represents LAI value frequencies in a 0.5 (m^2/m^2) interval.

Figure 4 shows the dominance map of global LAI products composited by reciprocal averages of MAE from GIMMS LAI4g (green), GIMMS LAI3g (red), and GLASS LAI (blue). Within each 2° × 2° grid, the color was determined by the LAI products with a lower MAE, i.e., a higher absolute LAI accuracy. An immediate observation from Figure 4 is that the absolute LAI accuracy of GIMMS LAI4g was significantly higher than others in most parts of the world. However, this advantage was relatively weak in the northern latitudes of the Eurasian continent (40°−60°). The

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

320 GIMMS LAI3g and GLASS LAI could show higher accuracy at rather random locations. We acknowledged that the number of Landsat LAI samples in certain 2° × 2° grids might not be sufficient for robust accuracy assessment, but that would not alter the overall outperformance of the GIMMS LAI4g.

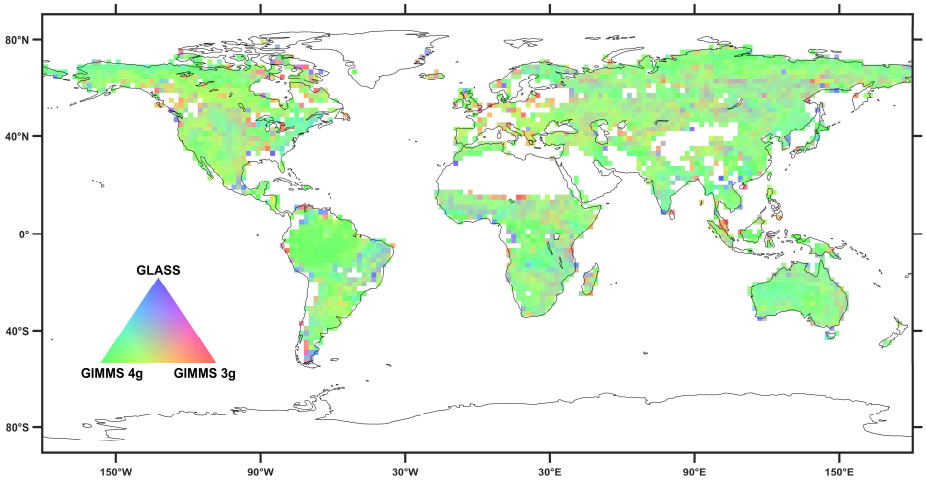
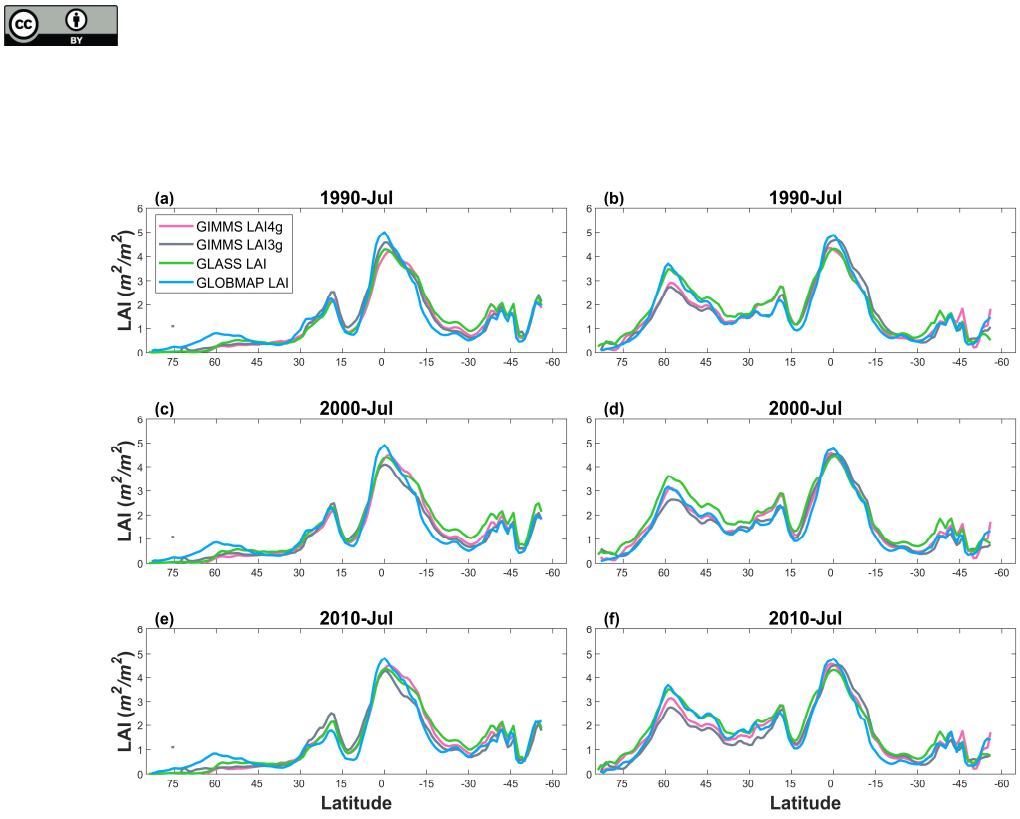


Figure 4. Dominance map of the GIMMS LAI3g, GIMMS LAI4g, and GLASS LAI based on their MAE. The map was drawn in 2° × 2°
325 grids whose colors were composed of reciprocal averages of MAE from the GIMMS LAI4g (green), GIMMS LAI3g (red), and GLASS LAI (blue). A greener grid, for example, indicates that the GIMMS LAI4g has a lower MAE (or a higher absolute LAI accuracy).

Figure 5 showcases the spatially averaged LAI along latitude in January and July of the years 1990, 2000, and 2010 for GIMMS LAI4g, GIMMS LAI3g, GLASS LAI, and GLOBMAP LAI, respectively. The four LAI products were overall
330 consistent. The GIMMS LAI4g and GIMMS LAI3g had lower values at northern middle latitudes (35°N – 65°N) in July (Figure 5b; Figure 5d; Figure 5f). Also in July, the GLOBMAP LAI and GLASS LAI in the Northern Hemisphere maintained good consistency for the years 1990 and 2010 (Figure 5b; Figure 5f), but the GLOBMAP LAI was systematically lower than GLASS LAI for the year 2000 (Figure 5d).

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023



335 **Figure 5.** Inter-comparison of spatially averaged LAI along latitude between the GIMMS LAI4g, GIMMS LAI3g, GLASS LAI, and GLOBMAP LAI in January and July of the years 1990, 2000, and 2010. The spatial average was calculated at an interval of 1°.

4.4 Temporal consistency analysis

Figure 6 shows the variations of LAI bias in EBF for the GIMMS LAI4g and other three LAI products. The LAI bias during different NOAA satellite missions was distinguished. The GIMMS LAI4g demonstrated an outstanding temporal
340 consistency with minimum bias variations (Figure 6a), indicating an efficient removal of satellite orbital drift and sensor degradation effects. LAI bias significantly fluctuated in the GIMMS LAI3g, GLASS LAI, and GLOBMAP LAI with different patterns. The GIMMS LAI3g relied on AVHRR data only and its bias varied with NOAA missions. The evident AVHRR degradation after the year 2012, as argued by Wang et al. (2022) can be also observed in our results (Figure 6b). The GLASS LAI and GLOBMAP LAI used different data sources before (AVHRR) and after (MODIS) the year 2000. For the GLASS LAI,
345 bias variations before 2000 were much larger than those after 2000 (Figure 6c). The reason is likely that the data quality from MODIS is better than AVHRR. For the GLOBMAP LAI, however, bias variations remained large for all periods of NOAA missions (Figure 6d).

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

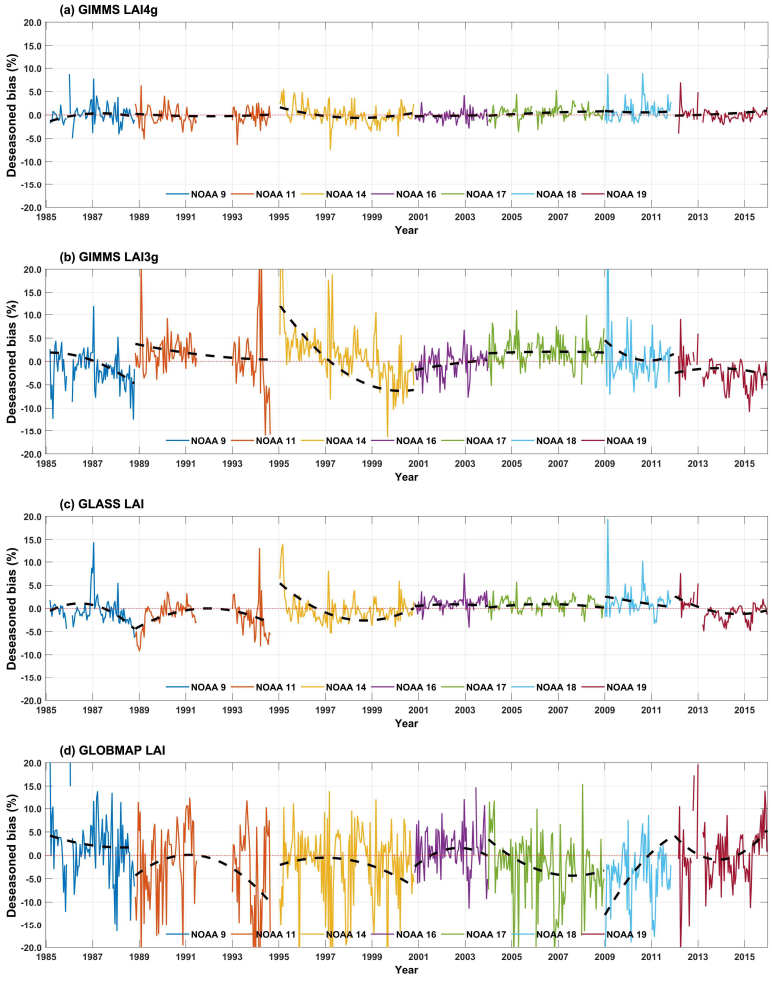


Figure 6. Temporal variations of LAI bias% in EBF for (a) the GIMMS LAI4g, (b) GIMMS LAI3g, (c) GLASS LAI, and (d) GLOBMAP LAI. The black dash line represents the interannual trend extracted by the EEMD method. Values from different NOAA satellite missions were distinguished with colors.

Figure 7 shows the GIMMS LAI4g time series before (thin black line) and after (bold colored line) data consolidation from 1982 to 2020 for different vegetation biome types. The GIMMS LAI4g shared the same footprint with MODIS LAI after the year 2004. Before consolidation, there was a systematic deviation between the GIMMS LAI4g and MODIS LAI products for all vegetation biomes. The pixel-wise linear fusion method has successfully matched the GIMMS LAI4g time series with

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

MODIS LAI, eliminating abnormal shifts in vegetation phonologies. The temporal variations of the GIMMS LAI4g after consolidation were self-consistent in all periods. All these results demonstrated that the consolidation of GIMMS LAI4g and MODIS LAI products was reliable.

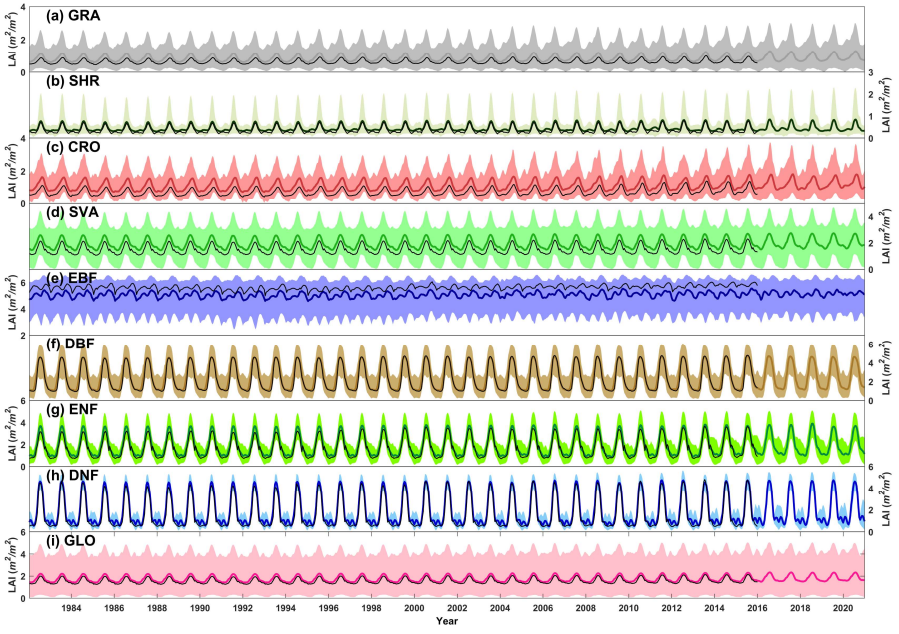


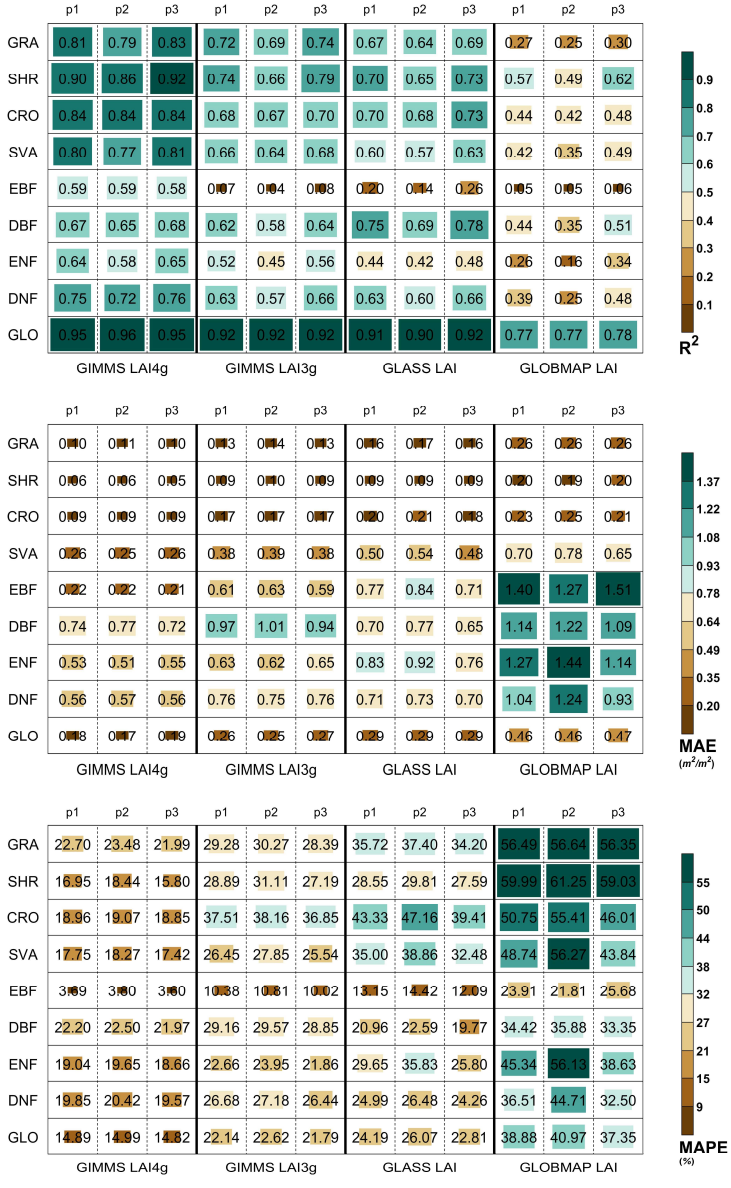
Figure 7. Temporal variations of the GIMMS LAI4g for different vegetation biome types during 1982−2020. GLO represents the global vegetation biome. The bold colored line represents the LAI average of GIMMSLAI4g after data consolidation, with shadow covering the value range between 10% and 90% quantiles. The thin black line represents the LAI average of GIMMSLAI4g before consolidation. It should be noted that the GIMMS LAI4g after consolidation shared the same footprint with the Reprocessed MODIS LAI after the year 2004.

Figure 8 shows the LAI accuracies in three periods, i.e., 1984−2015 (p1), 1984−2000 (p2), and 2001−2015 (p3), for the four global LAI products. The results show good temporal consistency for the GIMMS LAI4g and GIMMS LAI3g. Their accuracy differences between p2 and p3 (i.e., 1984−2000 and 2001−2015) were minimum for most vegetation biomes. In particular, the global vegetation biome shows constant R^2 values (GIMMS LAI4g: 0.96 (p2) vs. 0.95 (p3); GIMMS LAI3g: 0.92 vs. 0.92) (Figure 7a) and a small difference in MAE (GIMMS LAI4g: 0.17 m^2/m^2 vs. 0.19 m^2/m^2; GIMMS LAI3g: 0.25 m^2/m^2 vs. 0.27 m^2/m^2) and MAPE (GIMMS LAI4g: 14.99% vs. 14.82%; GIMMS LAI3g: 22.62% vs. 21.79%). The temporal consistency of GLOBMAP LAI in different periods was relatively low. For the GLASS LAI and GLOBMAP LAI that used different data sources before (AVHRR) and after (MODIS) the year 2000, data quality after 2000 was higher than that before

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

Open Access Earth System Science Data Discussions

2000 because of the improvement in satellite sensors. The GIMMS LAI4g product used Landsat LAI samples that covered the whole period from 1984 to 2015. This consistency in LAI reference data resulted in a minimum difference between periods.



https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

Figure 8. Temporal consistencies between different periods for the global LAI products. The global LAI products include GIMMS LAI4g, GIMMS LAI3g, GLASS LAI, and GLOBMAP LAI). The periods are 1984−2015 (p1), 1984−2000 (p2), and 2001−2015 (p3). The consistencies were evaluated at the biome level using R^2 (a), MAE (b), and MAPE (c) calculated based on Landsat LAI samples. GLO represents the global vegetation biome.

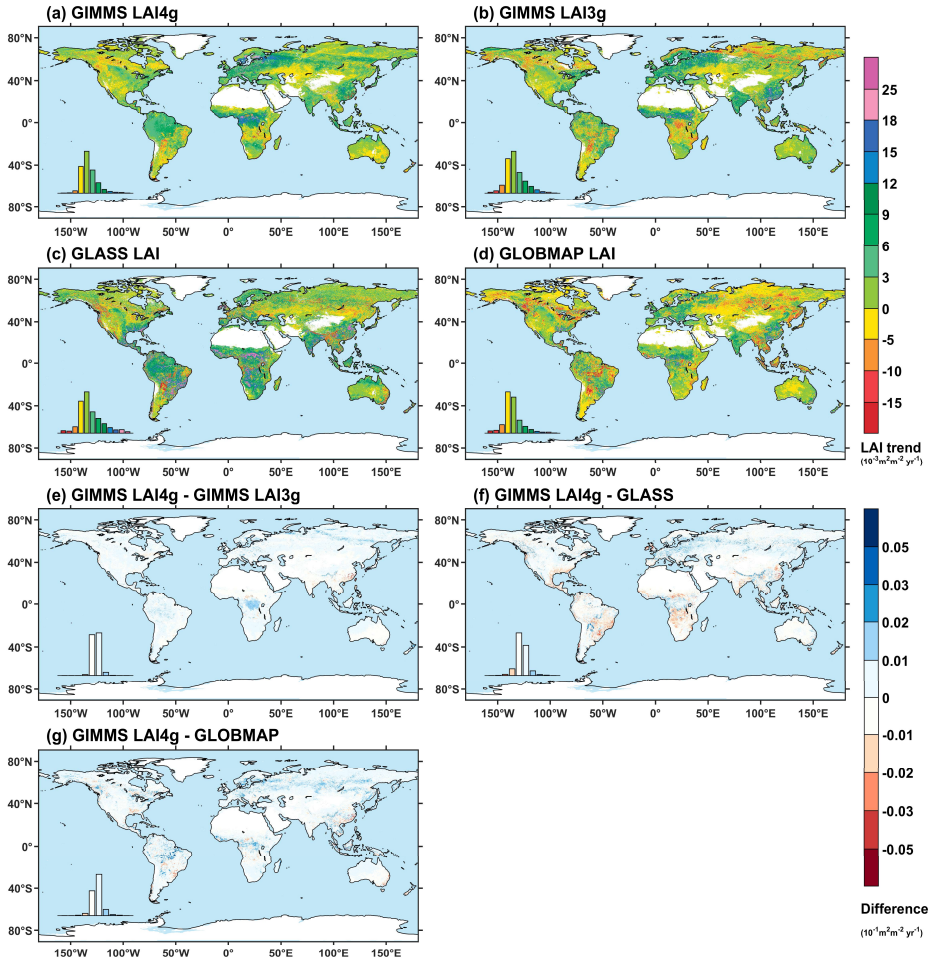
4.5 LAI trend analysis

Figure 9a to Figure 9d show the slope maps of the LAI time series from the GIMMS LAI4g, GIMMS LAI3g, GLASS LAI, and GLOBMAP in the period of 1982−2015. Figure 9e to Figure 9g shows the slope differences between the GIMMS LAI4g and the other three LAI products. In general, the GIMMS LAI4g, GIMMS LAI3g, and GLASS LAI showed a similar spatial pattern that agreed on the greening trend in global hotspot areas such as China and India. The GIMMS LAI4g demonstrated a more significant greening trend in the high-latitude regions of northern Europe and Asia.

Figure 10a shows the annual average LAI trends for different vegetation biomes in the four LAI products. The GIMMS LAI3g, GIMMSLAI4g, and GLASS LAI products presented a similar greening trend for the global vegetation biome, with a slope value of 2.06×10^{-2} $m^2m^{-2}yr^{-1}$, 1.22×10^{-2} $m^2m^{-2}yr^{-1}$, and 3.81×10^{-2} $m^2m^{-2}yr^{-1}$, respectively. The greening trend in the GLOBMAP LAI product was significantly lower, with a value of 0.05×10^{-2} $m^2m^{-2}yr^{-2}$. Except for ENF and DNF, the slope values of most biomes in the GIMMS LAI4g were in-between the maximum and minimum values of the other three LAI products.

Figure 10b shows the annual LAI variations of the four LAI products for different vegetation biomes. The GIMMS LAI3g and GLASS LAI showed remarkable differences in trend before and after the year 2000. Their LAI values significantly increased before 2000 but turned stagnant after 2000. The GIMMS LAI4g demonstrated continuous greening trends around the year 2000. This phenomenon was also observed in EBF. The GIMMS LAI3g had opposite trends to the GIMMS LAI4g after the year 2000, which may be caused by AVHRR sensor degradation (Wang et al., 2022), as MODIS LAI also showed a greening trend (Wang et al., 2022; Jiang et al., 2017) in this period.

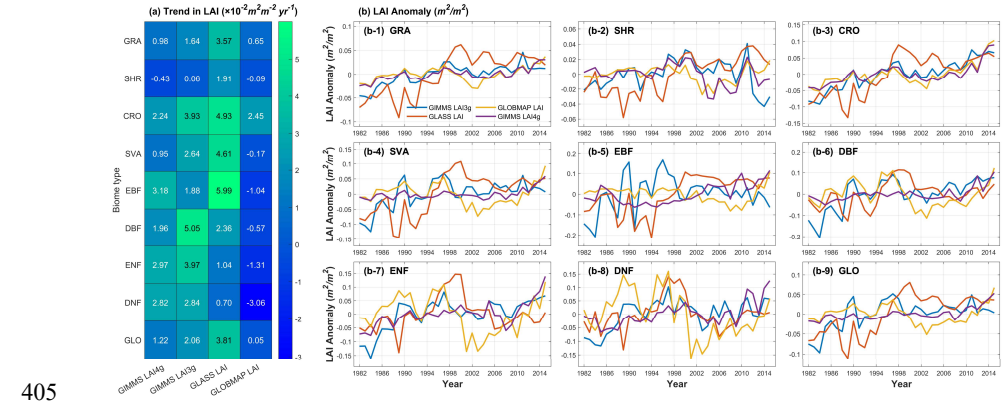
https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023



400

Figure 9. Global maps of LAI trends and their differences between the global LAI products during 1982−2015. The LAI products include GIMMS LAI4g (a), GIMMS LAI3g (b), GLASS LAI (c), and GLOBMAP LAI (d). The trend was calculated as the slope of linearly fitted LAI time series. (e) to (g) show the slope differences between the GIMMS LAI4g and the other three LAI products.

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023



405 **Figure 10.** Variations of annual LAI anomaly of different vegetation biomes in the global LAI products during 1982−2015. The LAI products include GIMMS LAI4g, GIMMS LAI3g, GLASS LAI, and GLOBMAP LAI. (a) shows the slope values of the annual LAI and (b) shows the annual LAI time series. GLO represents the global vegetation biome.

5. Discussion

410 5.1 Improvements over other long-term global LAI products

The remote sensing data source and LAI reference data are critical inputs for the accurate derivation of long-term and large-scale LAI products. The first improvement in this study was the use of a more reliable remote sensing product, i.e., the PKU GIMMS NDVI product. Some global LAI products (e.g., GLASS LAI and GLOBMAP LAI) directly used remote sensing surface reflectance as the data source (Ma and Liang, 2022; Kang et al., 2021; Xiao et al., 2016), but some others (e.g., GIMMS 415 LAI3g) argued that NDVI could be more robust against the terrain, atmospheric conditions, and BRDF effects (Zhu et al., 2013; Pinzon and Tucker, 2014; Zeng et al., 2022). More importantly, the PKU GIMMS NDVI product essentially addresses the issue related to the NOAA satellite orbital drift and AVHRR sensor degradation (Li et al., in review), which had widely existed in current long-term global LAI products (Zhu et al., 2013) (Figure 6) since the AHVRR data were the only data source before the late 1990s that provided spatiotemporal observations over the globe. Our GIMMS LAI4g thereby was also free from 420 this issue (Figure 6). The second improvement in our LAI product was the use of massive and high-quality Landsat LAI samples. In current long-term global LAI products (e.g., GLASS LAI, GLOBMAP LAI, and GIMMS LAI3g), the LAI reference data were either ground measurements which were spatially and temporally insufficient or LAI values derived from advanced sensors unavailable before the year 2000 (Zhu et al., 2013; Chen et al., 2019a). In the LAI model generalization, uncertainties could be hard to determine over locations and dates when the LAI reference data were basically absent. The

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

Landsat LAI samples used in this study had a large number (3.6 million), a long time series (1984−2015), and global coverage (Figure 1). These two improvements in this study guaranteed that our GIMMS LAI4g product is more spatiotemporally consistent, as demonstrated in our results (sections 4.3 and 4.4).

Besides, incorporating other explanatory variables, including spatial, temporal, and satellite-based information, further improved the robustness of LAI models. Specifically, the role of spatial information (longitude and latitude) and temporal information (month) has been underscored in explaining global LAI variations (Figure 2 and Table 1). In this study, individual BNPP models were developed for vegetation biomes. We chose not to include the vegetation biome type as an explanatory variable (GIMMS LAI3g did) because the values of the vegetation biome type are deterministic rather than continuous (like month).

Compared to its predecessor (GIMMS LAI3g; 1982−2016) which relied on AVHRR data only, our GIMMS LAI4g (1982−2020) provides up-to-date LAI data by consolidating with the MODIS LAI (the Reprocessed MODIS LAI product). This extension of temporal coverage could help interpret recent global vegetation dynamics. Two other LAI products, namely the GLASS LAI (1982−2018) and GLOBMAP LAI (1982−2020), also incorporated MODIS data (reflectance). However, they did not explicitly calibrate systematic deviations between AVHRR and MODIS data (Liu et al., 2012; Xiao et al., 2014). Our study employed a pixel-wise linear fusion method (Mao et al., 2012) to match the GIMMS LAI4g with the MODIS LAI product. The results showed an excellent consistency between the GIMMS LAI4g (after consolidation) and MODIS LAI.

5.2 Potential applications of the GIMMS LAI4g product

With an explicit physical meaning, LAI was proposed to be more accurate in characterizing vegetation dynamics than spectral indices such as NDVI and Enhanced Vegetation Index (EVI) (Zhang et al., 2004; Verger et al., 2016). Our results demonstrated that the GIMMS LAI4g product could be more spatiotemporally consistent and reliable than other long-term global LAI products. One important role of the GIMMS LAI4g is to mitigate the disagreements between current global LAI products and gain robust knowledge about long-term vegetation changes. For the past 40 years, the long-term analysis based on the global LAI products has shown an overall greening trend in most vegetated areas. However, significant variations existed between different LAI products at the regional scale (Wang et al., 2022; Jiang et al., 2017). In the evergreen broadleaf forests of Africa, for instance, the GIMMS LAI3g exhibited a decreasing trend from the year 2000 while the MODIS LAI exhibited an increasing trend (Wang et al., 2022). The GIMMS LAI4g provides an opportunity to better understand the spatial pattern of vegetation greening (or browning) and its drivers (Zhu et al., 2016; Piao et al., 2015; Chen et al., 2019a).

LAI is also a popular proxy for many important ecosystem attributes and functions, such as carbon stock and sink (Chen et al., 2019b), nutrition cycle (Pierce et al., 1994), and evapotranspiration (Wang et al., 2014). It serves as a fundamental parameter in many ecosystem models (Boussetta et al., 2013; Boussetta et al., 2015; Chen et al., 2015), earth system models (Mahowald et al., 2016), and climate models (Boussetta et al., 2013; Boussetta et al., 2015). The GIMMS LAI4g is expected

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

to benefit the development of these models and provide a powerful data basis for a more accurate and reliable land surface characterization.

5.3 Uncertainty source of GIMMS LAI4g product

The PKU GIMMS NDVI product and the Landsat LAI samples comprise this study's primary sources of uncertainty. Despite the efforts by Zha et al. (in preparation), the number of Landsat LAI samples was small in certain regions, e.g., the northern high latitudes and tropical areas. This was attributed to the low solar altitude angle, polar night phenomenon, and climate conditions such as frequent clouds and rains at the time of Landsat observation. The PKU GIMMS NDVI product suffers from the same issues as it was also derived from Landsat samples (Li et al., in review). Although both Li et al. (in review) and the current study have used MODIS data as compensation, the relative lack of Landsat samples in the northern high latitudes and tropical areas may still lower the robustness of models in the generation of GIMMS LAI4g to a certain extent.

Other sources of uncertainty could be from the BPNN model structure in which more explanatory variables such as temperature and precipitation could be incorporated, the static vegetation biome map which could be replaced by annual maps if available, and the Reprocessed MODIS LAI product. It should be noted that, however, these uncertainties also existed in other LAI products and this study has tried its best to mitigate their influence on the GIMMS LAI4g.

6. Conclusions

This study developed a new generation of the GIMMS LAI product (GIMMS LAI4g, 1982−2020) based on Back Propagation Neural Network (BPNN) models and a pixel-wise consolidation method. The GIMMS LAI4g was featured by the use of the PKU GIMMS NDVI product and the massive high-quality Landsat LAI samples. The recently published PKU GIMMS NDVI efficiently removed the effects of NOAA orbital drift and AVHRR sensor degradation, which has been a critical issue in existing LAI products. The high-quality global Landsat LAI samples, with a total number of 3.6 million and temporal coverage of 1984−2015, facilitated the creation of spatiotemporally consistent BPNN models. The spatiotemporally consistent GIMMS LAI4g product covers a time span of 1982 to 2020, with a spatial resolution of 1/12° and a temporal resolution of half-month. It can potentially provide strong data support for long-term vegetation monitoring and model development with high accuracy and reliability, as shown below:

- Evaluated by the Landsat LAI samples, the GIMMS LAI4g product (R^2=0.95, MAE=0.18 m^2/m^2, MAPE=15%) was overall more accurate than the mainstream global LAI products, including the GIMMS LAI3g (R^2=0.92, MAE=0.26 m^2/m^2, MAPE=22%), GLASS LAI (R^2=0.91, MAE=0.29 m^2/m^2, MAPE=24%), and GLOBMAP LAI (R^2=0.77,

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

MAE=0.46 m^2/m^2, MAPE=39%). Its accuracy meets the target proposed by the Global Climate Observation System (GCOS).

- The GIMMS LAI4g outperformed the other LAI products in most regions of the globe and most vegetation biomes (R^2: 0.59 to 0.90; MAE: 0.06 m^2/m^2 to 0.74 m^2/m^2; MAPE: 4% to 23%) except the deciduous broadleaf forests.
- The GIMMS LAI4g product removed the effects of NOAA orbital drift and AVHRR sensor degradation, which can be observed in other LAI products.
- The GIMMS LAI4g after consolidation with the Reprocessed MODIS LAI was more temporally consistent between the three periods of 1984−2015, 1984−2000, and 2001−2015 than other LAI products. It could also more reasonably depict global vegetation trends (greening or browning).

Appendix A

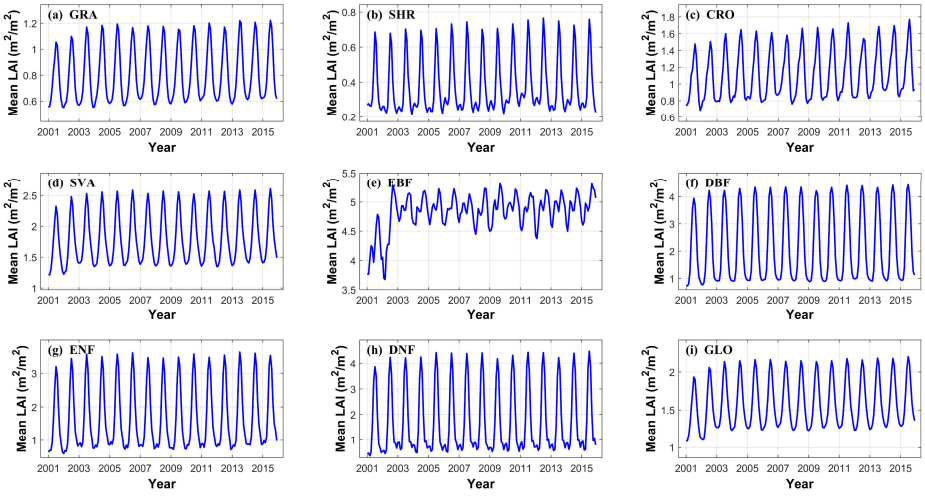


Figure A1. Annual LAI averages from the Reprocessed MODIS LAI product (version 6) for different vegetation biome types. The vegetation biome types use the third classification scheme in the MODIS Land-Cover Type product (MCD12Q1, version 6) (see section 2.4). GLO represents the global vegetation biome. This figure demonstrates unexpected low LAI values for evergreen broadleaf forests (EBF) between 2000 and 2003.

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

Appendix B

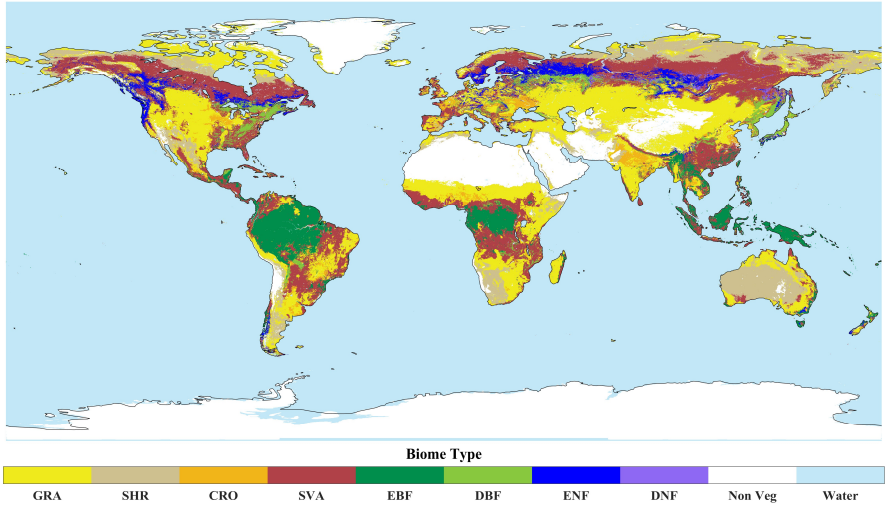


Figure B1. The global land cover map derived from the MODIS Land-Cover Type product (MCD12Q1, version 6). The third classification scheme in MCD12Q1 was adopted. Each grid (1/12°) in the map was labeled as the most frequent land cover type between 2001 and 2019.

Data availability

The spatiotemporally consistent global dataset of the GIMMS Leaf Area Index (GIMMS LAI4g) generated in this study is openly available at https://doi.org/10.5281/zenodo.7649108 (Cao et al., 2023). It covers the whole global vegetation area at half-month temporal resolution and 1/12° spatial resolution from 1982 to 2020. It is available in Geographic Lat/Lon projection and TIFF format.

Author contributions. ZZ conceptualized and supervised the project. ZZ, ML, SC, and JZ designed the workflow of methodology to product the dataset. ML, JZ, WZ, ZD, JC, YZ, and YC conducted the work in data acquisition and processing. ZZ, ML, SC, and JZ performed data analysis. SC, ML, and ZZ prepared the manuscript. SC and ZZ reviewed and edited the draft. All authors contributed to the interpretation of the results.

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this study.

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

Open Access Earth System Science Data Discussions

Acknowledgments. We would like to thank NASA for providing MODIS data products, Zhiqiang Xiao for providing GLASS LAI data, Yang Liu and Ronggao Liu for providing GLOBMAP LAI data, and Hua Yuan for providing reprocessed MODIS LAI product. We gratefully acknowledge the Landsat data support from USGS and Google Earth Engine (https://earthengine.google.com/). We are also grateful to the anonymous reviewers for their constructive comments and suggestions. The updated algorithm with the reviewers' suggestions has been used for generating the GIMMS LAI4g product.

Financial support. This work was supported by the National Natural Science Foundation of China (42271104, 41901122), the Shenzhen Fundamental Research Program (GXWD20201231165807007-20200814213435001), and the Shenzhen Science and Technology Program (JCYJ20220531093201004).

References

Alkama, R., Forzieri, G., Duveiller, G., Grassi, G., Liang, S., and Cescatti, A.: Vegetation-based climate mitigation in a warmer and greener World, Nat. Commun., 13, 606, https://doi.org/10.1038/s41467-022-28305-9, 2022.

Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Niño, F., Weiss, M., Samain, O., Roujean, J. L., and Leroy, M.: LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION: Part 1: Principles of the algorithm, Remote Sens. Environ., 110, 275-286, https://doi.org/10.1016/j.rse.2007.02.018, 2007.

Basheer, I. A. and Hajmeer, M.: Artificial neural networks: fundamentals, computing, design, and application, J. Microbiol. Methods, 43, 3-31, https://doi.org/10.1016/s0167-7012(00)00201-3, 2000.

Boussetta, S., Balsamo, G., Beljaars, A., Kral, T., and Jarlan, L.: Impact of a satellite-derived leaf area index monthly climatology in a global numerical weather prediction model, Int. J. Remote Sens., 34, 3520-3542, https://doi.org/10.1080/01431161.2012.716543, 2013.

Boussetta, S., Balsamo, G., Dutra, E., Beljaars, A., and Albergel, C.: Assimilation of surface albedo and vegetation states from satellite observations and their impact on numerical weather prediction, Remote Sens. Environ., 163, 111-126, https://doi.org/10.1016/j.rse.2015.03.009, 2015.

Broge, N. H. and Leblanc, E.: Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density, Remote Sens. Environ., 76, 156-172, https://doi.org/10.1016/s0034-4257(00)00197-8, 2001.

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

Cao, S., Li, M., Zhu, Z., Zha, J., Zhao, W., Duanmu, Z., Chen, J., Zheng, Y., and Chen, Y.: Spatiotemporally consistent global dataset of the GIMMS Leaf Area Index (GIMMS LAI4g) from 1982 to 2020 (V1.0), zenodo [dataset], https://doi.org/10.5281/zenodo.7649108, 2023.

Chen, C., Park, T., Wang, X. H., Piao, S. L., Xu, B. D., Chaturvedi, R. K., Fuchs, R., Brovkin, V., Ciais, P., Fensholt, R., Tommervik, H., Bala, G., Zhu, Z. C., Nemani, R. R., and Myneni, R. B.: China and India lead in greening of the world through land-use management, Nat. Sustainability, 2, 122-129, https://doi.org/10.1038/s41893-019-0220-7, 2019a.

Chen, J. M., Ju, W., Ciais, P., Viovy, N., Liu, R., Liu, Y., and Lu, X.: Vegetation structural change since 1981 significantly enhanced the terrestrial carbon sink, Nat. Commun., 10, 4259, https://doi.org/10.1038/s41467-019-12257-8, 2019b.

Chen, M., Willgoose, G. R., and Saco, P. M.: Investigating the impact of leaf area index temporal variability on soil moisture predictions using remote sensing vegetation data, J. Hydrol., 522, 274-284, https://doi.org/10.1016/j.jhydrol.2014.12.027, 2015.

Claverie, M., Matthews, J. L., Vermote, E. F., and Justice, C. O.: A 30+ year AVHRR LAI and FAPAR climate data record: algorithm description and validation, Remote Sens., 8, 263, https://doi.org/10.3390/rs8030263, 2016.

de Wit, A., Duveiller, G., and Defourny, P.: Estimating regional winter wheat yield with WOFOST through the assimilation of green area index retrieved from MODIS observations, Agric. For. Meteorol., 164, 39-52, https://doi.org/10.1016/j.agrformet.2012.04.011, 2012.

Dente, L., Satalino, G., Mattia, F., and Rinaldi, M.: Assimilation of leaf area index derived from ASAR and MERIS data into CERES-Wheat model to map wheat yield, Remote Sens. Environ., 112, 1395-1407, https://doi.org/10.1016/j.rse.2007.05.023, 2008.

Eyring, V., Gillett, N. P., Achutarao, K., Barimalala, R., Barreiro Parrillo, M., Bellouin, N., Cassou, C., Durack, P., Kosaka, Y., McGregor, S., Min, S.-K., Morgenstern, O., and Sun, Y.: Human Influence on the Climate System, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B. (eds.)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 423−552, https://doi.org/10.1017/9781009157896.005, 2021.

Fang, H., Baret, F., Plummer, S., and Schaepman-Strub, G.: An overview of global Leaf Area Index (LAI): methods, products, validation, and applications, Rev. Geophys., 57, 739-799, https://doi.org/10.1029/2018RG000608, 2019.

Friedl, M., Sulla-Menashe, D.: MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006, NASA EOSDIS Land Processes DAAC [dataset], 10.5067/MODIS/MCD12Q1.006, 2019.

Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., and Schaaf, C.: Global land cover mapping from MODIS: algorithms and early results, Remote Sens. Environ., 83, 287-302, https://doi.org/10.1016/s0034-4257(02)00078-0, 2002.

GCOS: Systematic observation requirements for satellite-based products for climate, 2011 update, 2011.

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

Helder, D., Thome, K. J., Mishra, N., Chander, G., Xiong, X. X., Angal, A., and Choi, T.: Absolute radiometric calibration of Landsat using a pseudo invariant calibration site, IEEE Trans. Geosci. Remote Sens., 51, 1360-1369, https://doi.org/10.1109/tgrs.2013.2243738, 2013.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. L. C., Shih, H. H., Zheng, Q. N., Yen, N. C., Tung, C. C., and Liu, H. H.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. A, 454, 903-995, https://doi.org/10.1098/rspa.1998.0193, 1998.

Jahan, N. and Gan, T. Y.: Modelling the vegetation-climate relationship in a boreal mixedwood forest of Alberta using normalized difference and enhanced vegetation indices, Int. J. Remote Sens., 32, 313-335, https://doi.org/10.1080/01431160903464146, 2011.

Jiang, C., Ryu, Y., Fang, H., Myneni, R., Claverie, M., and Zhu, Z. J. G. C. B.: Inconsistencies of interannual variability and trends in long-term satellite leaf area index products, Global Change Biol., 23, 4133-4146, https://doi.org/10.1111/gcb.13787, 2017.

Kang, Y., Ozdogan, M., Gao, F., Anderson, M. C., White, W. A., Yang, Y., Yang, Y., and Erickson, T. A.: A data-driven approach to estimate leaf area index for Landsat images over the contiguous US, Remote Sens. Environ., 258, 112383, https://doi.org/10.1016/j.rse.2021.112383, 2021.

Kimura, R., Okada, S., Miura, H., and Kamichika, M.: Relationships among the leaf area index, moisture availability, and spectral reflectance in an upland rice field, Agric. Water Manage., 69, 83-100, https://doi.org/10.1016/j.agwat.2004.04.009, 2004.

Li, M., Cao, S., and Zhu, Z.: Spatiotemporally consistent global dataset of the GIMMS Normalized Difference Vegetation Index (PKU GIMMS NDVI) from 1982 to 2020, Earth Syst. Sci. Data, in review. [see *review asset*]

Liu, Y., Liu, R., and Chen, J. M.: Retrospective retrieval of long-term consistent global leaf area index (1981-2011) from combined AVHRR and MODIS data, J. Geophys. Res.: Biogeosci., 117, https://doi.org/10.1029/2012jg002084, 2012.

Ma, H. and Liang, S.: Development of the GLASS 250-m leaf area index product (version 6) from MODIS data using the bidirectional LSTM deep learning model, Remote Sens. Environ., 273, 112985, https://doi.org/10.1016/j.rse.2022.112985, 2022.

Mahowald, N., Lo, F., Zheng, Y., Harrison, L., Funk, C., Lombardozzi, D., and Goodale, C.: Projections of leaf area index in earth system models, Earth Syst. Dyn., 7, 211-229, https://doi.org/10.5194/esd-7-211-2016, 2016.

Mao, D., Wang, Z., Luo, L., and Ren, C.: Integrating AVHRR and MODIS data to monitor NDVI changes and their relationships with climatic parameters in Northeast China, Int. J. Appl. Earth Obs. Geoinf., 18, 528-536, https://doi.org/10.1016/j.jag.2011.10.007, 2012.

Myneni, R., Knyazikhin, Y., Park, T: MOD15A2H MODIS/Terra Leaf Area Index/FPAR 8-Day L4 Global 500m SIN Grid V006, NASA EOSDIS Land Processes DAAC [dataset], 10.5067/MODIS/MOD15A2H.006, 2015a.

Myneni, R., Knyazikhin, Y., Park, T.: MCD15A2H MODIS/Terra+Aqua Leaf Area Index/FPAR 8-day L4 Global 500m SIN Grid V006, NASA EOSDIS Land Processes DAAC [dataset], 10.5067/MODIS/MCD15A2H.006, 2015b.

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R., and Running, S. W.: Global products of vegetation leaf

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

area and fraction absorbed PAR from year one of MODIS data, Remote Sens. Environ., 83, 214-231, https://doi.org/10.1016/S0034-4257(02)00074-3, 2002.

Panda, S. S., Ames, D. P., and Panigrahi, S.: Application of Vegetation Indices for Agricultural Crop Yield Prediction Using Neural Network Techniques, Remote Sens., 2, 673-696, https://doi.org/10.3390/rs2030673, 2010.

Piao, S., Yin, G., Tan, J., Cheng, L., Huang, M., Li, Y., Liu, R., Mao, J., Myneni, R. B., Peng, S., Poulter, B., Shi, X., Xiao, Z., Zeng, N., Zeng, Z., and Wang, Y.: Detection and attribution of vegetation greening trend in China over the last 30 years, Global Change Biol., 21, 1601-1609, https://doi.org/10.1111/gcb.12795, 2015.

Pierce, L. L., Running, S. W., and Walker, J.: Regional-scale relationships of leaf area index to specific leaf area and leaf nitrogen content, Ecol. Appl., 4, 313-321, https://doi.org/10.2307/1941936, 1994.

Pinzon, J. E. and Tucker, C. J.: A non-stationary 1981-2012 AVHRR NDVI3g yime series, Remote Sens., 6, 6929-6960, https://doi.org/10.3390/rs6086929, 2014.

Tucker, C. J., Pinzon, J. E., Brown, M. E., Slayback, D. A., Pak, E. W., Mahoney, R., Vermote, E. F., and El Saleous, N.: An extended AVHRR 8‐km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data, Int. J. Remote Sens., 26, 4485-4498, https://doi.org/10.1080/01431160500168686, 2005.

Valderrama-Landeros, L. H., España-Boquera, M. L., and Baret, F.: Deforestation in Michoacan, Mexico, from CYCLOPES-LAI time series (2000–2006), IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 9, 5398-5405, https://doi.org/10.1109/JSTARS.2016.2597742, 2016.

Verger, A., Filella, I., Baret, F., and Penuelas, J.: Vegetation baseline phenology from kilometric global LAI satellite products, Remote Sens. Environ., 178, 1-14, https://doi.org/10.1016/j.rse.2016.02.057, 2016.

Wang, L., Good, S. P., and Caylor, K. K.: Global synthesis of vegetation control on evapotranspiration partitioning, Geophys. Res. Lett., 41, 6753-6757, https://doi.org/10.1002/2014GL061439, 2014.

Wang, Z., Wang, H., Wang, T., Wang, L., Liu, X., Zheng, K., and Huang, X.: Large discrepancies of global greening: Indication of multi-source remote sensing data, Global Ecol. Conserv., 34, e02016, https://doi.org/10.1016/j.gecco.2022.e02016, 2022.

Xiao, Z., Liang, S., and Jiang, B.: Evaluation of four long time-series global leaf area index products, Agric. For. Meteorol., 246, 218-230, https://doi.org/10.1016/j.agrformet.2017.06.016, 2017.

Xiao, Z. Q., Liang, S. L., Wang, J. D., Xiang, Y., Zhao, X., and Song, J. L.: Long-time-series global land surface satellite leaf area index product derived from MODIS and AVHRR surface reflectance, IEEE Trans. Geosci. Remote Sens., 54, 5301-5318, https://doi.org/10.1109/tgrs.2016.2560522, 2016.

Xiao, Z. Q., Liang, S. L., Wang, J. D., Chen, P., Yin, X. J., Zhang, L. Q., and Song, J. L.: Use of general regression neural networks for generating the GLASS leaf area index product from time-series MODIS surface reflectance, IEEE Trans. Geosci. Remote Sens., 52, 209-223, https://doi.org/10.1109/tgrs.2013.2237780, 2014.

Yan, K., Park, T., Yan, G., Liu, Z., Yang, B., Chen, C., Nemani, R. R., Knyazikhin, Y., and Myneni, R. B.: Evaluation of MODIS LAI/FPAR product collection 6. Part 2: validation and intercomparison, Remote Sens., 8, 460, https://doi.org/10.3390/rs8060460, 2016.

https://doi.org/10.5194/essd-2023-68
Preprint. Discussion started: 24 February 2023

Yan, K., Park, T., Chen, C., Xu, B., Song, W., Yang, B., Zeng, Y., Liu, Z., Yan, G., Knyazikhin, Y., and Myneni, R. B.: Generating global products of LAI and FPAR from SNPP-VIIRS data: theoretical background and implementation, IEEE Trans. Geosci. Remote Sens., 56, 2119-2137, https://doi.org/10.1109/TGRS.2017.2775247, 2018.

Yuan, H., Dai, Y., Xiao, Z., Ji, D., and Shangguan, W.: Reprocessing the MODIS Leaf Area Index products for land surface and climate modelling, Remote Sens. Environ.,115, 1171-1187, https://doi.org/10.1016/j.rse.2011.01.001, 2011.

Zeng, Y. L., Hao, D. L., Huete, A., Dechant, B., Berry, J., Chen, J. M., Joiner, J., Frankenberg, C., Bond-Lamberty, B., Ryu, Y., Xiao, J. F., Asrar, G. R., and Chen, M.: Optical vegetation indices for monitoring terrestrial ecosystems globally, Nat. Rev. Earth Environ., 3, 477-493, https://doi.org/10.1038/s43017-022-00298-5, 2022.

Zha, J.: The inversion and application of global high and moderate resolution leaf area index validation dataset, M.S. thesis, Peking University, Beijing, 72 pp., 2022.

Zha, J., Li, M., Zhu, Z., Cao, S., Zhang, Y., Zhao, W., and Chen, Y.: Spatiotemporally consistent global Landsat leaf area index validation dataset, in preparation. [see *review asset*]

Zhang, G. Q., Patuwo, B. E., and Hu, M. Y.: Forecasting with artificial neural networks: The state of the art, Int. J. Remote Sens., 14, 35-62, https://doi.org/10.1016/s0169-2070(97)00044-7, 1998.

Zhang, P., Anderson, B., and Barlow, M.: Climate-related vegetation characteristics derived from Moderate Resolution Imaging Spectroradiometer (MODIS) leaf area index and normalized difference vegetation index, J. Geophys. Res.: Atmos., 109, https://doi.org/10.1029/2004jd004720, 2004.

Zhu, Z., Bi, J., Pan, Y., Ganguly, S., Anav, A., Xu, L., Samanta, A., Piao, S., Nemani, R. R., and Myneni, R. B.: Global data sets of vegetation Leaf Area Index (LAI)3g and Fraction of Photosynthetically Active Radiation (FPAR)3g derived from Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index (NDVI3g) for the period 1981 to 2011, Remote Sens., 5, 927-948, https://doi.org/10.3390/rs5020927, 2013.

Zhu, Z., Piao, S., Myneni, R. B., Huang, M., Zeng, Z., Canadell, J. G., Ciais, P., Sitch, S., Friedlingstein, P., Arneth, A., Cao, C., Cheng, L., Kato, E., Koven, C., Li, Y., Lian, X., Liu, Y., Liu, R., Mao, J., Pan, Y., Peng, S., Peñuelas, J., Poulter, B., Pugh, T. A. M., Stocker, B. D., Viovy, N., Wang, X., Wang, Y., Xiao, Z., Yang, H., Zaehle, S., and Zeng, N.: Greening of the Earth and its drivers, Nat. Clim. Change, 6, 791-795, https://doi.org/10.1038/nclimate3004, 2016.