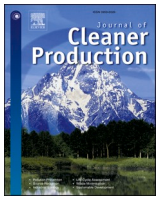
Journal of Cleaner Production 509 (2025) 145616

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro



Disentangling gross primary productivity drivers of forested areas in China and its climate zones from 1990 to 2018

Chenxi Zhu [a], Guojie Wang [a], Yuhao Shao [b], Wen Dai [b], Qiang Liu [a], Shuangao Wang [c,d], Ana Cristina Costa [c], Pedro Cabral [a,c,*]

[a] School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science & Technology, Nanjing, 210044, China
[b] School of Geographical Sciences, Nanjing University of Information Science & Technology, Nanjing, China
[c] NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312, Lisboa, Portugal
[d] Information Center, Beijing Academy of Science and Technology, Beijing, 100089, China

ARTICLE INFO

Handling Editor: Liu Yu

Keywords:
Spatiotemporal clusters
Climate change
Machine learning
Forest fragmentation index
Chatterjee correlation

ABSTRACT

Gross primary productivity plays a critical role in global carbon balance. However, quantifying the effects of different drivers still constitutes a challenge due to the different modeling techniques and data used. This study employs spatiotemporal analysis, machine learning, and statistical approaches to measure the significance of forest gross primary productivity drivers in China and its climate zones from 1990 to 2018. The results show that the annual average forest gross primary productivity in China was 914.74 gC m^{-2} y^{-1} during the study period and showed a significantly increasing trend ($p < 0.01$) at a rate of 4.09 gC m^{-2} y^{-1}. Forest gross primary productivity had a southeast-northwest downward spatiotemporal trend with significantly different distributions within the six climate zones, except in the arid and semi-arid zones. A Random Forest model did better than an eXtreme Gradient Boosting model when 10 explanatory variables were used. These variables included the novel forest fragmentation index and climate zones, which helped explain the effects of forest structure and climate characteristics of the climate zones better. The most important forest gross primary productivity drivers in China were mean annual temperature (26.2 %), mean annual precipitation (18.6 %), solar radiation (11 %), forest fragmentation index (8.8 %), and elevation (8.1 %). Using Chatterjee's correlation coefficient, this study provides, for each climate zone, its unique signature regarding the order and importance of the drivers of forest gross primary productivity. This study helps us understand what factors affect forest gross primary productivity in China and its climate zones better by showing how they work using machine learning. These findings may help China reach its carbon neutrality goals.

1. Introduction

Gross Primary Productivity (GPP) is the total quantity of carbon produced by photosynthesis in plants (Beer et al., 2010). Terrestrial GPP constitutes the greatest flow component of the global carbon budget, and forests are essential to the terrestrial carbon cycle (Zhang et al., 2022). Globally, GPP is predicted to increase until the year 2100 in high-latitude regions and decrease in equatorial regions under future climate change (Lu et al., 2024). Precisely characterizing the spatiotemporal changes of GPP and its drivers is essential for comprehending and identifying potential interventions in the biogeochemical dynamics of terrestrial ecosystems, carbon budget, and climate change (Prakash Sarkar et al., 2022). However, the direct quantification of terrestrial GPP can only be measured at the leaf level through photosynthesis measurements, making it a difficult task (Welp et al., 2011). Thus, indirect measurements of GPP have been the purpose of many studies using global dynamic vegetation models (Li et al., 2016) or ground observations with eddy covariance techniques at ecosystem scale (Pastorello et al., 2020). These observations have been successfully integrated with remote sensing (RS) for implementation of vegetation GPP monitoring (Ding et al., 2024). Nonetheless, ecological process models rely on model parameterization and climate forcing variables, and the simulation results may vary importantly (Liao et al., 2023). This problem may be efficiently tackled through the application of machine learning models in regression and prediction tasks using independent explanatory variables, i.e., drivers, to describe their relationship to GPP

* Corresponding author. School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science & Technology, Nanjing, 210044, China.
E-mail address: cabral@nuist.edu.cn (P. Cabral).

https://doi.org/10.1016/j.jclepro.2025.145616
Received 26 September 2024; Received in revised form 24 April 2025; Accepted 27 April 2025
Available online 29 April 2025
0959-6526/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

measurements (Gaber et al., 2024).

The drivers used in GPP modelling studies usually include climatic, topographic, and/or anthropogenic factors. The lack of precipitation has been considered one of the most important factors negatively affecting GPP (Hoover et al., 2021). Air temperature, which interacts with precipitation, is also associated with GPP (Yin et al., 2022). While temperature may enhance GPP by promoting photosynthesis and extending the growing season, an excessive increase in temperature may also lead to stress on plants, dehydration, and increased respiration, which can reduce GPP (Beer et al., 2010). Soil moisture enables plants to maintain healthy stomatal conductance, which facilitates the uptake of CO_2 for photosynthesis, thus supporting higher GPP (Beer et al., 2010). However, studies have demonstrated a higher importance of moisture variables in non-forests compared to forests (Chang et al., 2023). The main energy source for GPP, solar radiation, directly influences photosynthesis and is a significant factor in determining how much carbon plants absorb (Wang et al., 2022). Topography and human activity influence the spatial pattern of forest Net Primary Productivity (NPP), which varies across altitude and slope gradients (Chen et al., 2023). Slopes facing the sun receive more direct sunlight, which promotes higher rates of photosynthesis and GPP. Researchers have linked a decline in GPP to an increase in human activities like urbanization and agriculture (Guo et al., 2023). However, we still need to understand and clarify the role of diverse drivers in shaping the forest GPP to better model this phenomenon and define adaptation measures.

The abovementioned drivers, among others, can be successfully integrated in machine learning models to simulate GPP through the use of techniques such as Random Forests (RF) (Yao et al., 2018), Support Vector Machines (SVM) (Yujie et al., 2023), Boosted Regression Trees, Artificial Neural Networks (ANN) (Guo et al., 2023), Convolutional Neural Networks (CNN) (Yu et al., 2022), Long Short-Term Memory networks (LSTM) (Guo et al., 2023), and Deep Neural Networks (DNN) (Hu et al., 2024). Seemingly, machine learning models are increasingly being utilized to investigate complex phenomena that vary greatly over time and space, such as GPP (Guo et al., 2023). However, while there are a great number of studies that have used machine learning methods to estimate GPP, results depend on the characteristics of the study areas, selected variables, data spatiotemporal resolution and acquisition characteristics, and model selection and parameterization (Reichstein et al., 2019). Thus, there is still a need to further expand the knowledge of GPP estimation with machine learning methods (Prakash Sarkar et al., 2022).

Regarding the geographical area of application, it is possible to find machine learning modelling studies of GPP applied globally (Beer et al., 2010) and nation/region-wise, such as in China (Lu et al., 2024), Europe (Jung et al., 2007), Australia (Prakash Sarkar et al., 2022), the United States (Beaver et al., 2024), and other locations. Specifically for China, many studies have been carried out aiming to understand the different impacts of factors on GPP, including the use of techniques such as the two-leaf light use efficiency (TL-LUE) model (Zhang et al., 2022), geodetectors (Guo et al., 2023), nonlinear Granger causality, and eddy covariance (Zhang et al., 2024). Examples of machine learning methods include the study by Yao et al. (2018) that applied machine learning to develop a GPP dataset based on eddy flux measurements, Lu et al. (2023) that used a deep learning model to improve the accuracy of GPP estimation, Zhu et al. (2023) that mapped the Chinese annual GPP of terrestrial ecosystems using eddy covariance measurements and machine learning, and Chang et al. (2023) that used random forest regression to predict site-level GPP. Focusing on maize GPP, Liao et al. (2024) quantified irrigation effects using a machine learning framework using the eXtreme Gradient Boosting (XGBoost) model and SHapley Additive exPlanations. We conclude that, despite the increasing number of studies about GPP and related drivers using diverse methods in China, the distinction between the climate change and land cover change roles on GPP still requires further investigation (Zhang et al., 2022). The existing research is mostly applied to specific parts and/or provinces of China using diverse modeling approaches and variables and, consequently, yields different results regarding the most important drivers, making the knowledge about GPP changes on the Chinese mainland unclear. Additionally, none of the previous studies focused on forested areas using information about the forest structure nor fully considered the differences within the Chinese climate zones, which play an important role in shaping GPP.

This study aims to disentangle the role of different drivers in forest GPP in China and its climate zones. Besides several climate- and topographic-related drivers commonly used in GPP modeling, it incorporates two new variables. One of the variables is the novel forest fragmentation index (FFI), which is a measure of forest disturbance proposed by Ma et al. (2023). Forest structure is known to impact the forest GPP (Morreale et al., 2021). Since FFI is mostly human-induced, the inclusion of this driver, together with nighttime lights (Zhang et al., 2024), acts as proxies of human activities potentially impacting forest GPP. To better consider the importance of climate characteristics in the modelling procedure, this study introduces an additional variable, i.e., climate zones, which is a proxy for other climatic drivers that were not included in the modelling stage (e.g., atmospheric pressure, wind, etc.). Moreover, climate zones may approximate the spatial effects in forest GPP, such as spatial autocorrelation, which are not addressed in the machine learning models used in this study. Additionally, these models' evaluation metrics are biased if spatial autocorrelation is not accounted for (Meyer and Pebesma, 2022). Thus, using explicit spatial covariates, such as climate zones, is one of the most common approaches to deal with spatial autocorrelation in supervised machine learning regression (Jemeljanova et al., 2024). Specifically, this study aims to (i) describe and analyze the spatiotemporal trends of forest GPP between 1990 and 2018; (ii) identify a suitable machine learning approach to model forest GPP; and (iii) quantify the contribution of drivers of forest GPP in China and in its climate zones. This study provides novel information on the historical determinants of forest GPP in China and its climate zones, which can be useful for achieving carbon neutrality through informed planning.

2. Data and methods

2.1. Study area

The study area is mainland China, which spans more than 9.6 million km^2 (Fig. 1). Due to China's complex geography and unique climate, temperature and precipitation have important seasonal variations, mostly influenced by the East Asian monsoon. The average annual precipitation trends downward from the southeast to the northwest, and the average temperature exhibits a declining tendency from the north to the south. Using the Köppen climate classification (Kottek et al., 2006), China can be divided into six climate zones: arid, humid, temperate semi-humid, cold semi-humid, Qinghai-Tibet Plateau, and semi-arid.

The arid region's primary ecosystem is desert, characterized by extremely low precipitation, high evaporation, sparse vegetation, and low productivity (X. Wang et al., 2022). Super-xerophytic species, such as trees, subshrubs, semi-shrubs, shrubs, and herbaceous plants, dominate the vegetation in this region (X. Wang et al., 2022). The humid region's primary ecosystem is subtropical forests, thriving in a warm and humid climate with dense evergreen and mixed forests (Chen et al., 2024). These forests support a highly productive ecosystem. The main ecosystems in the temperate semi-humid zone are deciduous forests and agricultural landscapes, with vegetation primarily composed of subtropical evergreen and deciduous broadleaf forests (Yang et al., 2024). The eastern areas of this zone have highly developed agriculture, with wheat and maize serving as the primary crops. In the cold semi-humid zone, the dominant ecosystems are temperate forests, which feature deciduous broadleaf and mixed coniferous-broadleaf forests (Xiao-Ying et al., 2013). In addition, drier or transitional areas are home to cold-tolerant shrubs and herbaceous plants. Temperate grasslands, such

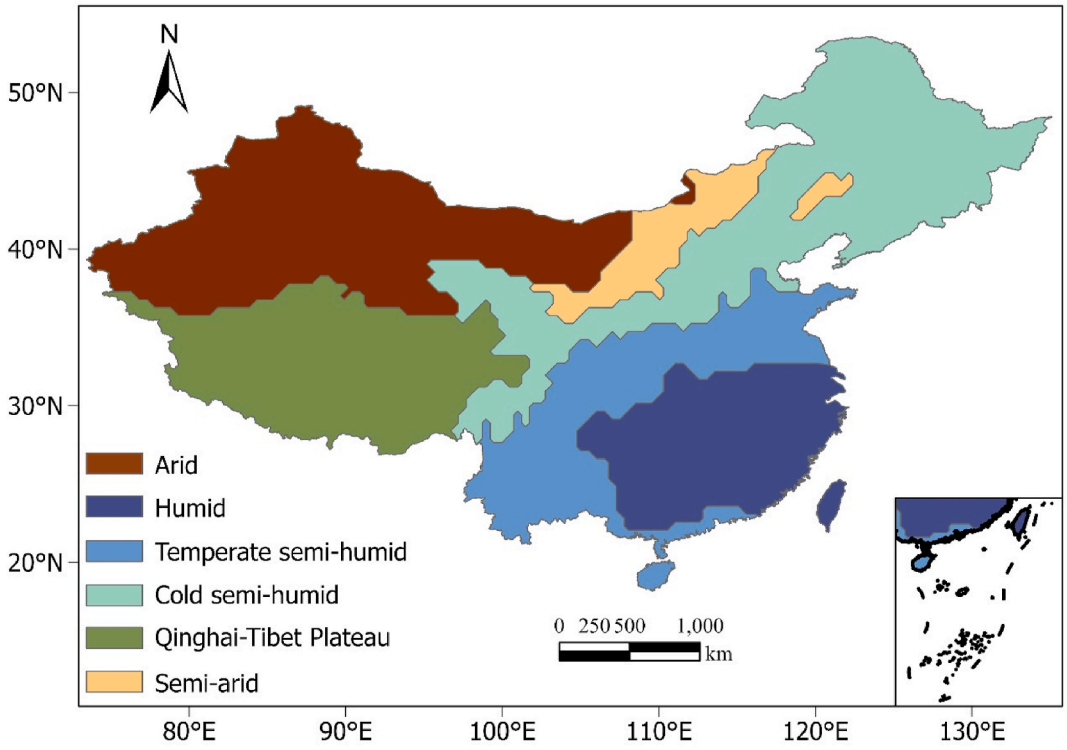


Fig. 1. Study area divided by the six climate zones according to Köppen climate classification (Kottek et al., 2006).

as meadow steppe and desert steppe, define the semi-arid zone, their vegetation well adapted to low precipitation and significant temperature fluctuations (Yin et al., 2016). Alpine meadow and alpine steppe ecosystems dominate the Qinghai-Tibet Plateau, their vegetation primarily composed of cold-resistant species like sedges, drought-tolerant grasses, sparse shrubs, and cushion plants (Hao et al., 2021). The extreme altitude and harsh climate of the plateau have uniquely adapted these plants.

2.2. Data

The data used in this study is described in Table 1. A long-term global GPP dataset, based on satellite near-infrared reflectance (NIRv) generated from 1982 to 2018 using the Advanced Very-High-Resolution Radiometer (AVHRR) and flux station observations and with a resolution of 0.05° per month, was used. This dataset captured accurately monthly (R^2 = 0.71) and annual (R^2 = 0.74) variations in GPP when compared to data from 104 flux sites (Wang et al., 2021). The data unit of GPP is gC m^{-2} d^{-1}.

Table 1
Data used in this study.

Dataset	Spatial resolution	Time span	Units	Source
Gross Primary Productivity	0.05°	1982–2018	gC m^{-2} d^{-1}	Wang et al. (2021)
Temperature	0.0083333°	1901–2022	°C	Peng et al. (2019)
Precipitation	0.1°	1901–2022	0.1 mm	Peng et al. (2019)
Downward shortwave radiation	0.1°	1979–2018	W m^{-2}	He et al. (2020)
Soil moisture	0.1°	1982–2020	m^3 m^{-3}	Zhang et al., 2024
Nighttime light	0.0083333°	1984–2020	Digital Number	(L. Zhang et al., 2024)
Land cover	30m	1990–2022	–	Yang and Huang (2021)
Digital Elevation Model	90m	2000	m	CGIAR-CSI (2022)

The temperature and precipitation datasets have the monthly average temperature and total precipitation data for China using a spatial resolution of 0.0083333° from the years 1901–2022 (Peng et al., 2019). The data units are °C and 0.1 mm, respectively. This data is based on the global 0.5° climate dataset created by the Climatic Research Unit (CRU) (https://crudata.uea.ac.uk/), and the global high-resolution climate dataset is produced by WorldClim (https://worldclim.org/) and was generated by a delta spatial downscaling scheme. i.e., a framework that uses low-spatial-resolution monthly time series and high-spatial-resolution reference climatology data for climate data downscaling in China (Peng et al., 2019). According to these authors, this dataset has better evaluations when compared to WorldClim and CRU data using accuracy assessment metrics like root mean square error, mean absolute error, and Nash-Sutcliffe efficiency coefficients.

The downward shortwave radiation was retrieved from the China Meteorological Forcing Dataset (CMFD), which is a gridded near-surface meteorological dataset produced since 1979. The dataset was created through fusion of remote sensing products, reanalysis datasets, and in-situ observation data obtained from weather stations (He et al., 2020). According to these authors, validation results using independent stations had higher quality when compared to other datasets like the Global Land Data Assimilation System (GLDAS). This dataset has a temporal resolution of 3 h and a spatial resolution of 0.1°.

The soil moisture dataset is gathered from the Simple Terrestrial Hydrosphere model version 2 (SiTHv2). It has a 0.1° spatial resolution and was obtained using hydrometeorological variables from reanalysis and satellite data (Zhang et al., 2024). According to these authors, using in-situ measurements and comparisons with other soil moisture products, this dataset showed a robust performance at multiple scales.

This study uses the annual Prolonged Artificial Nighttime-light DAtaset (PANDA) for China from 1984 to 2020. The dataset uses digital number (DN) values ranging from 0 to 63, which is a unitless scale. This scale represents the brightness intensities captured by the sensors,

with higher values indicating brighter light intensity observed at night. This dataset was produced using the Night-Time Light convolutional Long Short-Term Memory (NTLSTM) network and has a spatial resolution of 0.0083333° (L. Zhang et al., 2024). According to these authors, using quantitative and visual comparisons with other nighttime light datasets, this dataset offers superior temporal consistency and correlations with socioeconomic activities.

A land cover dataset containing nine land cover classes with a spatial resolution of 30m covering the period from 1990 to 2022 for China was used. This dataset was built using Landsat images with an annual temporal resolution (Yang and Huang, 2021). According to these authors, this dataset has 79.31 % overall accuracy when compared to visually interpreted samples, outperforming other reputed datasets.

In this study, the Digital Elevation Model (DEM) used has 90m spatial resolution and is an improved version from the National Geospatial Intelligence Agency (NGA) (https://www.nga.mil/), named Shuttle Radar Topography Mission (SRTM) 3-2. This is the fourth version of the SRTM data created by the International Center for Tropical Agriculture (CIAT) using an improved algorithm to fill SRTM 90 data holes. This dataset has a nominal absolute elevation accuracy of ±16m and an absolute plane accuracy of ±20m (CGIAR-CSI, 2022).

2.3. Methods

The methodological workflow comprises three stages and is described in Fig. 2. Details on the methods used are described in the following subsections.

2.3.1. Data pre-processing

A total of 10 drivers were selected to explain forest GPP: soil moisture, mean annual temperature, mean annual precipitation, downward solar radiation, elevation, slope, aspect, and nighttime lights. The drivers used in this study, apart from FFI and climate zones, are commonly used in GPP modelling studies (Chang et al., 2023). All the drivers are continuous variables, except climate zones, which is a categorical variable.

The land cover dataset was used to extract the forest class and calculate the FFI (Ma et al., 2023). Slope and aspect variables were calculated from the Digital Elevation Model (DEM) using ArcGIS Pro 3.3 software (https://www.esri.com/).

FFI measures the intensity of forest disturbance (Ma et al., 2023) and is an equally weighted normalized average of three landscape metrics (McGarigal et al., 2013): Edge Density (ED), Patch Density (PD), and Mean Patch Area (MPA, using 1-normalized MPA). The metrics are normalized to enable time series analysis using a grid of 5 km resolution. Its value varies between 0, no disturbance, and 1, high disturbance. This indicator was calculated using the 'landscapemetrics' package (Hesselbarth et al., 2019) available in R software (https://www.r-project.org/).

All the datasets were extracted from open data sources available in the National Tibetan Plateau Scientific Data Center (https://data.tpdc.ac.cn/). The overlapping period of the datasets from 1990 to 2018 was selected for this study. All the raster data was upscaled to 0.1° with a bilinear resample function and used a WGS 1984 World Mercator projection.

2.3.2. Spatiotemporal analysis of forest GPP and associated drivers

The trend of forest GPP and associated drivers was measured using the non-parametric Theil–Sen median method (Sen, 1968; Theil, 1950). Long-time series data trend analysis commonly uses this technique to identify patterns amidst outliers or non-normal data distributions (Theil, 1992). The Mann–Kendall (MK) test (Mann, 1945), which has reduced sensitivity to outliers and is independent from sample distributions, was used to evaluate the significance of the forest GPP trend and associated explanatory drivers. We conducted this analysis using the MATLAB software R2022b (https://www.mathworks.com/).

In the spatial distribution of trends, the values used for each variable were calculated based on annual values. Specifically, GPP and

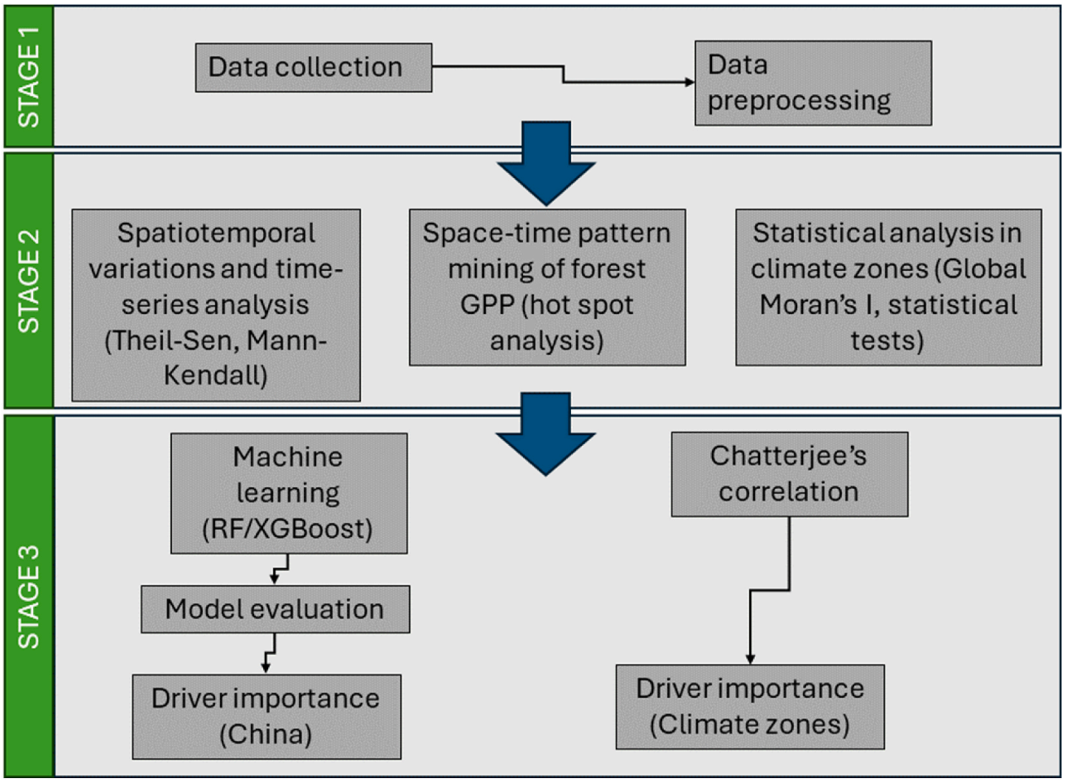


Fig. 2. Methodological workflow (Stage 1: Data collection and preprocessing; Stage 2: Spatiotemporal variations and analysis; Stage 3: Modelling and driver analysis).

precipitation (PRE) were represented as annual totals, while other variables, such as temperature and shortwave radiation, were represented as annual averages. These annual values were used to determine the spatial distribution of trends across the study region. For the time series analysis, the annual values of each variable (e.g., GPP annual total, PRE annual total, or the annual average for other variables) were averaged across all pixels within the study area for each year. This produced a single representative annual average value per year for the entire study area, providing insights into temporal variations at the regional scale.

Emerging Hot Spot Analysis (EHSA) facilitates the identification of patterns within the clustering points of forest GPP values within a space-time cube (STC). This type of analysis has been used in the context of agricultural climate suitability (Pan et al., 2023), urban heat island pattern analysis (Deng et al., 2023), hydroclimate extremes analysis (Palmer et al., 2024), and urban public safety (He et al., 2024), among other applications. The EHSA employs the Getis-Ord Gi* statistic (Getis and Ord, 1992), which uses time to assess a variable, in contrast to standard hot spot analysis, which only considers space (ESRI, 2023). We carried out the detection of spatiotemporal hot/cold spots in conjunction with the MK trend method (Mann, 1945). By using the z-score and p-value for each bin, the emerging hotspot analysis makes a unique classification for each site in the study area based on its space-time forest GPP pattern (Table 2). We conducted these analyses using ArcGIS Pro 3.3 (https://pro.arcgis.com).

The Global Moran's Index (I) test identifies statistically significant spatial autocorrelation in geographical processes (Moran, 1950). It involves evaluating the pattern of a set of features (i.e., points) and an associated attribute value (i.e., the forest GPP value) (Moran, 1948, 1950). The value for Global Moran's I ranges from −1 (dispersion of forest GPP) to 1 (clustering of forest GPP) when the numerator of the index is standardized by the variance. A value of 0 means that the distribution of forest GPP values is random. The Anderson-Darling test (Anderson and Darling, 1954) was applied to forest GPP values within each climate zone. This is one of the most powerful normality tests for asymmetric distributions (Yap and Sim, 2011). Further statistical analysis included non-parametric Kruskal-Wallis (Kruskal and Wallis, 1952) and Steel-Dwass (Dwass, 1960; Steel, 1960, 1961) tests to compare the distribution of forest GPP in each climate zone. All statistical tests were performed in JMP Pro 17, except the Global Moran's I test, which was computed in ArcGIS Pro 3.2.2. All conclusions are stated at the 5 % significance level.

Table 2
Description of spatiotemporal evolution patterns of forest GPP (ESRI, 2023).

Pattern name	Definition
No trend detected	Does not fall into any of the patterns defined below.
New hot (cold) spot	A location that is a statistically significant hot (cold) spot for the final time step but has never been before.
Consecutive hot (cold) spot	A location with a single uninterrupted run of at least two statistically significant hot (cold) spot bins in the final time-step intervals. The location has never been a statistically significant hot (cold) spot prior to the final hot (cold) spot run, and less than 90 % of all bins are statistically significant hot (cold) spots.
Intensifying hot (cold) spot	A location that has been a statistically significant hot (cold) spot for 90 % of the time-step intervals with a statistically significant increase in the clustering intensity overall.
Persistent hot (cold) spot	A location that has been a statistically significant hot (cold) spot for 90 % of the time-step intervals with no discernible trend in the clustering intensity over time.
Diminishing hot (cold) spot	A location that has been a statistically significant hot (cold) spot for 90 % of the time-step intervals with a statistically significant decrease in the clustering intensity overall.
Sporadic hot (cold) spot	A statistically significant hot (cold) spot for the final time-step interval with a history of also being an on-again and off-again hot (cold) spot. Less than 90 % of the time-step intervals have been statistically significant hot (cold) spots, and none of the time-step intervals have been statistically significant cold (hot) spots.
Oscillating hot (cold) spot	A statistically significant hot (cold) spot for the final time-step interval with a history of also being a statistically significant cold (hot) spot during a prior time step. Less than 90 % of the time-step intervals have been statistically significant hot (cold) spots.
Historical hot (cold) spot	The most recent time period is not a hot (cold) spot, but at least 90 % of the time-step intervals have been statistically significant hot (cold) spots.

2.3.3. Forest GPP modelling and drivers' correlation analysis

We implemented two different commonly used machine learning models to assess their ability to model mean forest GPP between 1990 and 2018 using the annual values averaged for each grid cell in each dataset of all variables, except the topographical variables, which remain constant over time, and the climate zones, which is a categorical variable. By using different models, we attempted to verify our analysis and guarantee the reliability of our conclusions by contrasting the outcomes from these two approaches.

Breiman et al. (2017) designed RF as an ensemble learning technique to enhance the classification and regression trees (CART) method. RF fits multiple decision tree classifiers on different subsamples of a dataset through averaging to reduce overfitting and improve predictive accuracy (Breiman, 2001). The algorithm uses bootstrap aggregation to carry out classification, producing many training subsets and trees, each one producing a classification outcome (Hunter et al., 2020). Using out-of-bag sampling and a replacement method, RF handles effectively dimensionality, multicollinearity, noise, outliers, and sample insufficiency (Hudak et al., 2020). Mean decrease impurity (MDI) values were used to assess factors importance in the RF ensemble models. MDI represents the cumulative reduction in variance caused by a feature (i.e., driving factor of forest GPP) across all splits in all trees in the forest (Ishwaran, 2015). At each split in a decision tree, the improvement in purity (i.e., reduction in variance in forest GPP) caused by the split is recorded for the selected feature. For each split, the reduction in variance is the difference between the variance of the parent node and the weighted variance of the (left and right) child nodes, where the weights are proportional to the number of samples in each child node. The MDI score for a feature, also known as factor importance, is the average of the variance reductions for all splits using that feature across all trees.

XGBoost was originally proposed by Chen and Guestrin (2016) to extend the gradient boosting framework where the tree structure consists of decision trees added sequentially to the model, and each tree aims to correct the errors of the previous ones. Initially, a simple model predicts the target variable, and subsequent trees are constructed to predict the residuals, which are the differences between the actual values and the predictions from the current model. These residuals represent the errors that the model needs to address. By iteratively fitting new trees to these residuals, XGBoost minimizes the overall loss function, thus improving the model's accuracy. Regularization techniques are applied to each tree to prevent overfitting, and the learning rate (shrinkage) ensures that each tree's contribution is scaled appropriately, allowing the model to converge smoothly. Bentéjac et al. (2021) explore the implementation aspects and detailed algorithmic contributions of XGBoost, demonstrating its versatility and accuracy performance.

Both machine learning models were implemented in ArcGIS Pro 3.3 using the Forest-based and Boosted Classification and Regression tool. The models were trained and evaluated using independent datasets (i.e., separate train-test folds). Before model fitting, 10 % of the total 53,314 points of the spatial file were randomly split as an independent validation set following model implementation recommendations in the ArcGIS tool. To evaluate the performance of the forest GPP models, three metrics were used, namely the coefficient of determination (R^2), the Mean Absolute Error (MAE), and the Root Mean Square Error (RMSE), using Eq. (1) to Eq. (3).

$$R^2 = \frac{\sum_{i=1}^{n}(P_i - \overline{O}_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O}_i)^2} \quad \text{(Eq.1)}$$

$$MAE = \frac{\sum_{i=1}^{n}|P_i - O_i|}{n} \quad \text{(Eq.2)}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \quad \text{(Eq.3)}$$

where P_i and O_i are the predicted and observed values for the GPP, respectively, and n is the sample size.

The R^2 specifies the ability of the independent variable to explain the changes in the dependent variable by the model. R^2 has a range of 0–1,

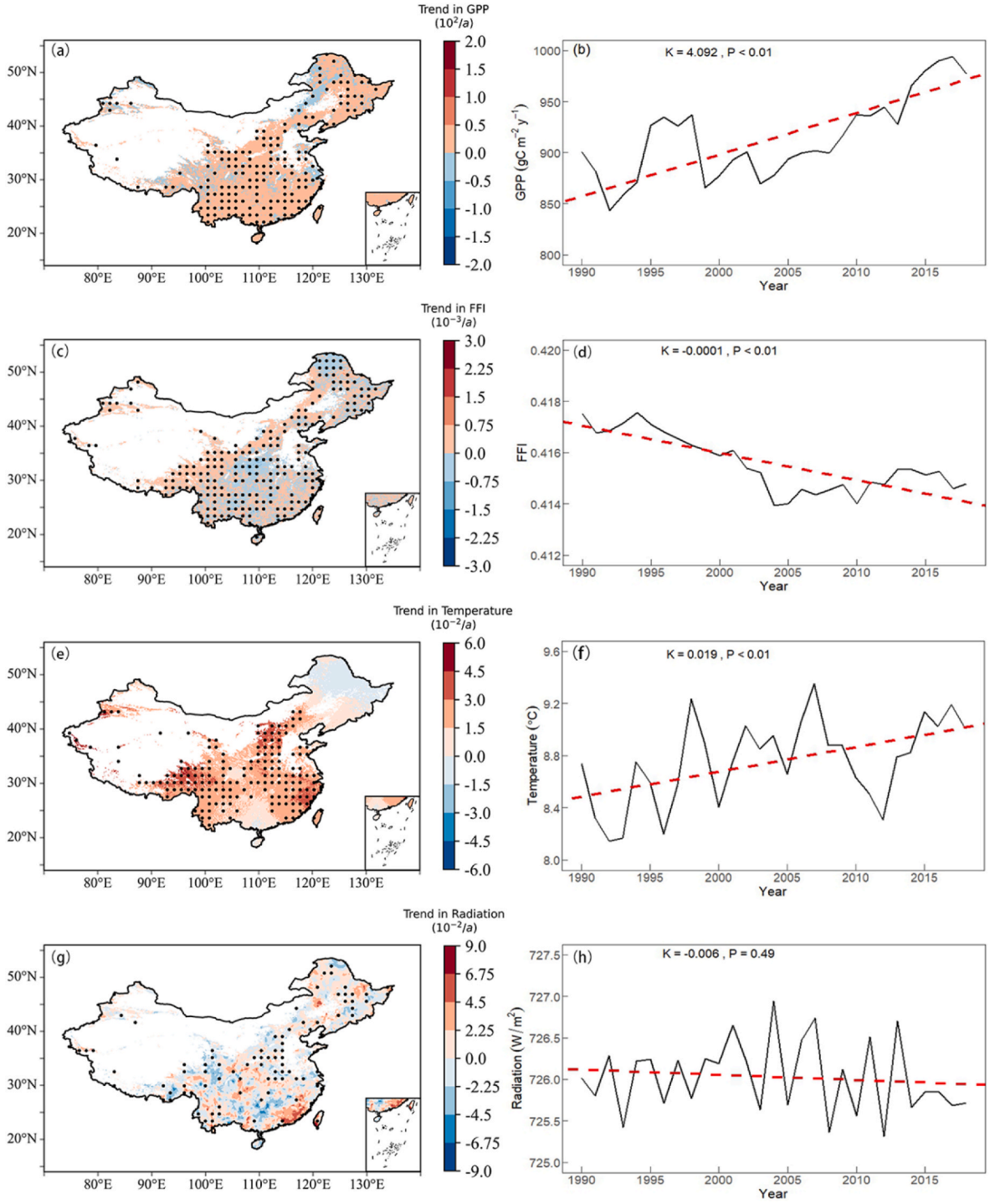


Fig. 3. Spatial patterns (left) and time series (right) of each variable in Chinese forest regions from 1990 to 2018. GPP (Gross Primary Productivity) (a, b), FFI (Forest Fragmentation Index) (c, d), temperature (e, f), radiation (g, h), soil moisture (i, j), precipitation (k, l), and nighttime light (m, n) are shown. Regions with significant trends ($P < 0.05$) are marked with black dots. The "10^x/a" above the color bar represents the annual trend of the variable, where "10^x" indicates the magnitude by which the values should be scaled. k and P denote the slope and p-value of each variable, respectively. The red dashed line represents the linear trend line.

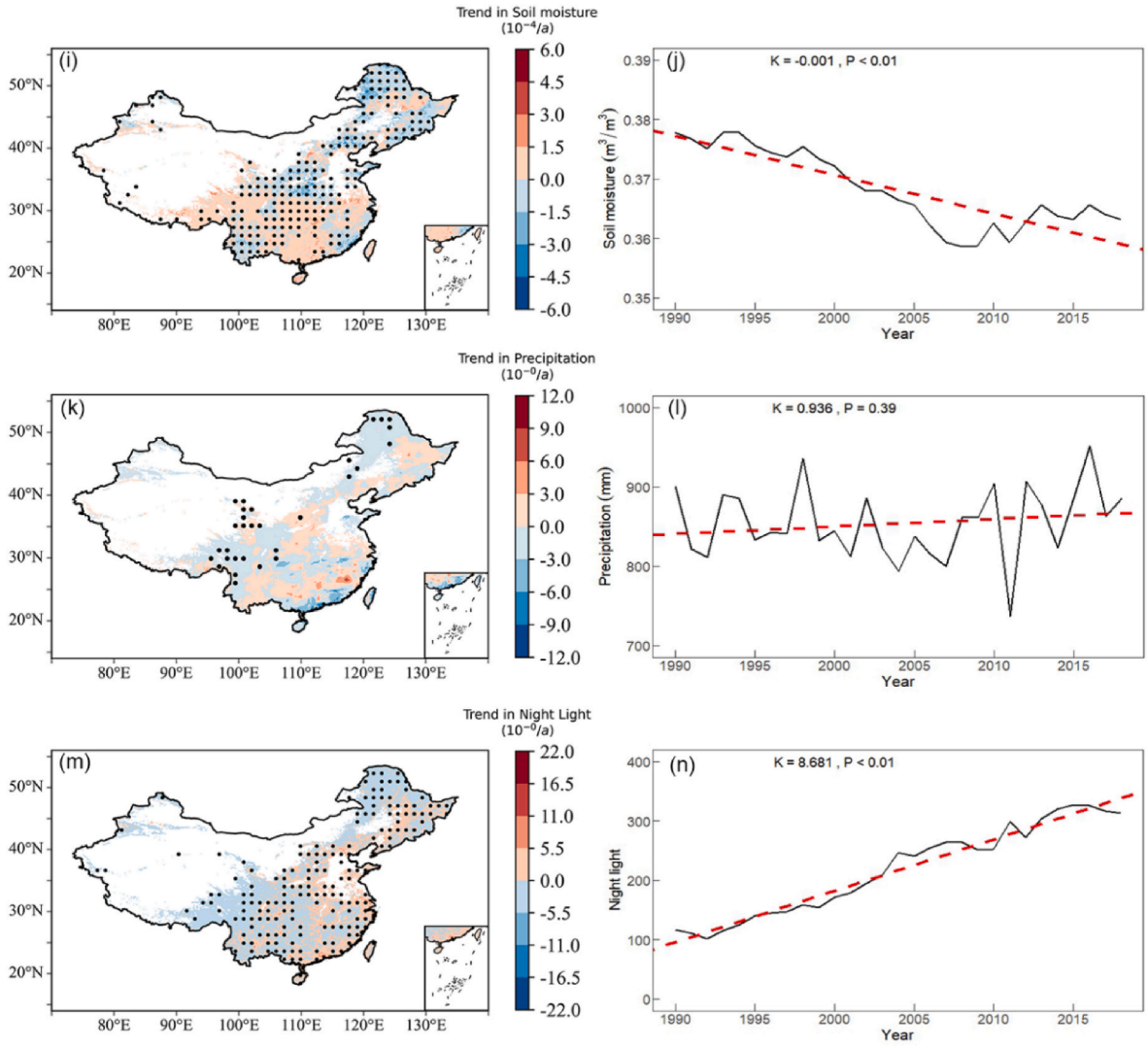


Fig. 3. (*continued*).

with a value near 1 denoting a good fit and a value near 0 denoting a bad fit. The average of the residuals in the dataset is captured by the mean absolute difference (MAE) between the expected and actual values. The standard deviation of the predicted errors, or RMSE, refers to the concentration of the data around the best fit line. Both the MAE and RMSE values are greater than or equal to zero. A high R^2 value combined with low MAE and RMSE values typically indicates an outstanding prediction performance.

To further investigate the strength and significance of the relationship between each driver (X) and forest GPP (Y) within each climate zone, the recent Chatterjee's correlation coefficient (Chatterjee, 2021) was computed. The ξ(X, Y) coefficient ranges from 0 to 1. Higher values indicate a stronger association. A value of one indicates a perfect association such that the Y variable is a measurable function of the X variable. A value of zero indicates no association. Chatterjee's correlation does not differentiate between positive and negative associations but focuses on the strength of the relationship, whether it is monotonic or non-monotonic (e.g., convex or concave). The p-value of the Chatterjee's nonparametric test for independence was estimated using 1000 permutations. The Chatterjee's correlation tests were obtained with the XICOR package (Chatterjee, 2021) in R 4.4.1. All conclusions are stated at the 5 % significance level.

3. Results

3.1. Spatiotemporal trends of forest GPP and its drivers

The mean annual forest GPP in China between 1990 and 2018 was 914.74 gC m^{-2} y^{-1}. During the study period, approximately 74 % of the forest area in China showed an increasing trend in GPP (with 63 % of the area showing significant increases, $P < 0.05$; Fig. 3a). However, 26 % of the forest area experienced a decreasing trend in GPP (with 34 % of the area showing significant decreases, $P < 0.05$; Fig. 3a), mainly concentrated in the middle and lower reaches of the Yangtze River, the Greater Khingan Range in the northeast, and the Hengduan Mountains. Overall, GPP increased at an average rate of 4.09 gC m^{-2} y^{-1} during the study period ($P < 0.01$; Fig. 3b). Notably, approximately 60 % of the forest area exhibited fragmentation trends, with a significant increase in FFI ($P < 0.05$; Fig. 3c). The spatial variation in FFI trends was largely opposite to those of GPP in most regions. In areas where FFI decreased significantly, GPP showed a significant increase, particularly in the Greater Khingan Range, the Hengduan Mountains, and the regions centered around the Qinling Mountains. The contrasting relationship between FFI and GPP implies a close relationship between human activities and changes in forest GPP in China. The rate of FFI reduction was estimated at 0.0001 per year ($P < 0.01$; Fig. 3d).

Additionally, forests across China have experienced a significant

warming trend, affecting 89 % of the area, with the most pronounced changes observed in the Hengduan Mountains, southeastern coastal regions, and the northeastern part of the Loess Plateau (significant in 60 % of the area). In contrast, only a few areas, such as Northeast China, exhibited a cooling trend, though these were not statistically significant (0 % of the area; Fig. 3e). The temporal analysis shows that, overall, the mean temperature rose significantly, with an annual increase of 0.019 °C ($P < 0.01$; Fig. 3f). The southern coast exhibited the most pronounced increasing trend in solar radiation (significant in 27 % of the area; Fig. 3g). In contrast, the central part of South China, Hainan Island, and areas north of the Hengduan Mountains displayed decreasing trends (significant in 26 % of the area; Fig. 3g). Across the entire study area, mean forest surface solar radiation declined slightly by 0.006 W m^{-2}, though this trend was not statistically significant ($P = 0.49$; Fig. 3h). Soil moisture exhibited significant geographic variation, increasing in the central part of South China (44 % of the area) while significantly decreasing in the north of the Qinling Mountains, the southwestern Yunnan-Guizhou Plateau, the southeastern coastal regions, and Northeast China (48 % of the area; Fig. 3i). Over the study period, soil moisture showed a significant overall decline at a rate of -0.001 m^3 m^{-3} per year ($P < 0.01$; Fig. 3j). The spatial distribution of precipitation and GPP trends varied, with both showing declines in the southern coastal regions, the Hengduan Mountains, and the Greater Khingan Mountains (Fig. 3k). Overall, total precipitation exhibited a slight upward trend, though this was not statistically significant ($P > 0.05$; Fig. 3l). The spatial distribution of trends in nighttime light was roughly opposite to that of GPP, particularly in the Yangtze River Delta, most parts of South China, and north of the Greater Khingan Range (Fig. 3m). However, overall, like GPP, it exhibited a significant increasing trend, with an annual growth rate of 8.681 and a total increase of 198 during the study period ($P < 0.01$; Fig. 3n). However, it showed a significant overall increasing trend, with an annual growth rate of 8.681 and a total increase of 198 during the study period ($P < 0.01$; Fig. 3n).

A ~50 km spatial resolution grid with the average of the 20 nearest-neighbors of forest GPP was used to calculate the emerging hot spots of forest GPP between 1990 and 2018 in China (Fig. 4). In the figure, it is possible to distinguish a northwest downward spatiotemporal trend in forest GPP, which is in conformity with China's terrain characteristics and climatic conditions. Most of the forest areas during the studied period do not exhibit any spatiotemporal pattern (29.3 %) in GPP. The analysis detected that 22.8 % of the forest area has an intensifying hot spot of forest GPP, especially in humid, tempered semi-humid and in the northern part of the cold semi-humid climate zones. This indicates that during the study period, the GPP values of these regions were significantly higher than those of other regions for at least 26 years, and the intensity of clustering was increasing. Persistent cold spots (14.6 %) were detected mostly in the Qinghai-Tibet Plateau and some in the northern parts of the arid and semi-arid climate zones. Sporadic hot spots (11 %) were found mostly in the humid and temperate semi-humid climate zones and in the northeast and southwest of the cold semi-humid zone, indicating that, except for 2018, these regions repeatedly appeared as hot spots throughout the historical period, with most of the time showing a significant increase, though in some periods the increase was not significant. Diminishing cold spots (9.5 %) are found in the arid, Qinghai-Tibet Plateau, cold semi-humid and semi-arid climate zones. These are mainly distributed in the northwest of arid, middle of semi-arid, and cold semi-humid areas, indicating that the forest GPP values in these areas are on a downward trend. Intensifying cold spots (5 %) were detected mostly in the Qinghai-Tibet Plateau, northwest of Arid, indicating that forest GPP values in these regions have always been low. Overall, the forest productivity in the humid, temperate semi-humid, and cold semi-humid climate zones exhibited an increasing trend during the study period, with relatively stable changes.

Results of the Global Moran's I test (Moran, 1950) show that forest GPP values exhibit a significant positive spatial autocorrelation in each climate zone, denoting clustering of forest GPP (Table S1). These results

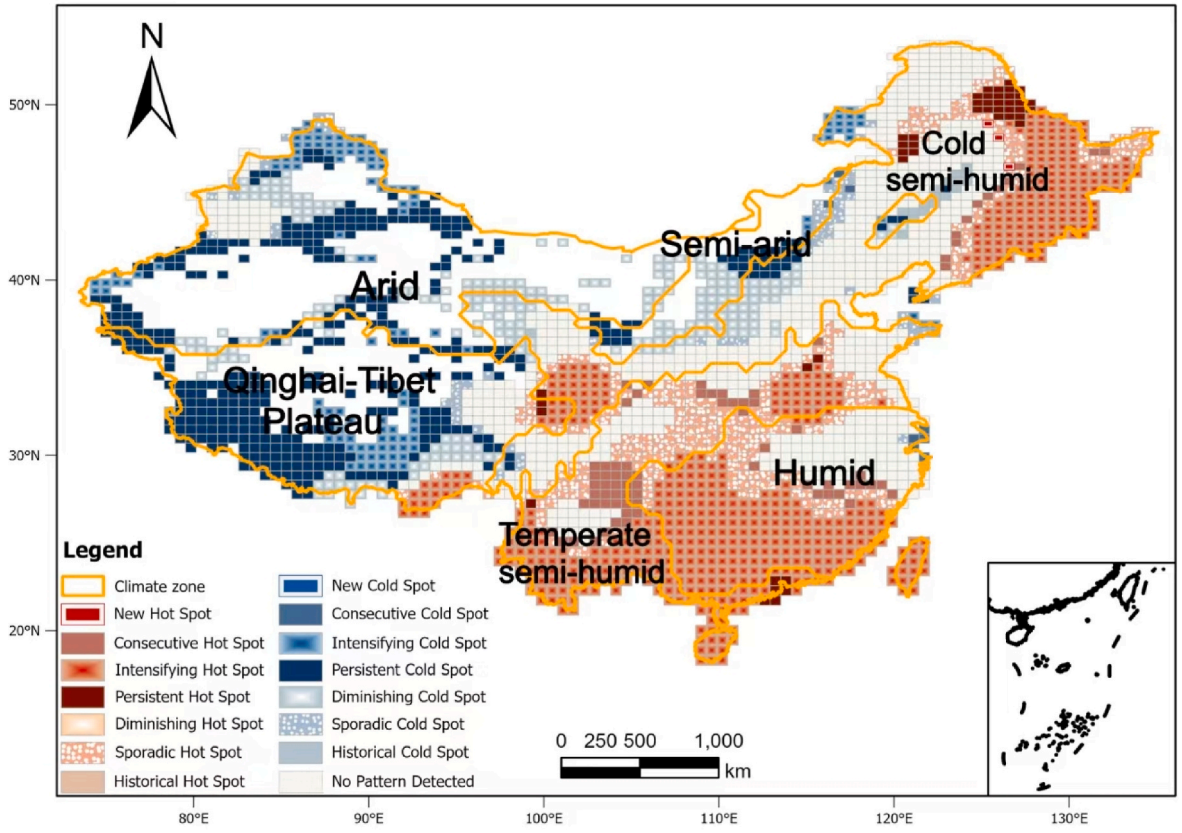


Fig. 4. Emerging spatial-temporal hot spots of forest GPP between 1990 and 2018 in mainland China using a spatial resolution of ~50 km.

were obtained using the Inverse Weighting scheme to assess each point's weight within the context of neighboring points. The Anderson-Darling test (Anderson and Darling, 1954) was applied to forest GPP values within each climate zone, and results show that there is not enough evidence that forest GPP data follows a Normal distribution in any of the climate zones (Table S2). The parametric one-way ANOVA technique is not appropriate to compare the means of forest GPP in the six climate zones because its normality assumption did not hold. Hence, the nonparametric Kruskal-Wallis test (Kruskal and Wallis, 1952) was used to assess if the location parameter of the forest GPP distribution is identical in all six climate zones. Results show that there is a significant difference in forest GPP across the six climate zones ($\chi 2 = 9279.499$; df $= 5$; p-value <0.001). A nonparametric post-hoc test was then applied, namely the Steel-Dwass test (Dwass, 1960; Steel, 1960, 1961). Results show that all forest GPP distributions are significantly different from any of the other distributions, except for arid (6) and semi-arid (3) zones, where forest GPP does not significantly differ (Table S3).

3.2. Machine learning model of forest GPP

Both RF and XGBoost models used 9 continuous variables (SM, TEMP, PRE, RAD, DEM, SLOPE, ASP, FFI, and LIGHT) and 1 categorical variable (climate zones). A total of 1500 trees with a minimum leaf size of 3 were used to train the RF model. The RF model obtained an R^2 of 0.73, an MAE of 208.13, and an RMSE of 325.97. On the other hand, XGBoost obtained an R^2 of 0.69, an MAE of 225.34, and an RMSE of 343.92. Since the RF outperformed XGBoost according to these metrics, the RF model was retained for further analysis of forest GPP.

3.3. Contribution of drivers to forest GPP

According to the RF model, the importance of the variables is ranked as follows (Fig. 5): temperature (26.2 %), precipitation (18.6 %), solar radiation (11 %), FFI (8.8 %), elevation (8.1 %), soil moisture (7.6 %), climate zone (6.8 %), nighttime lights (5.7 %), slope (4.6 %), and aspect (2.7 %). The contribution (%) corresponds to the percentage of the total sum of MDI scores. The values in the X axis are the MDI scores.

As shown in Fig. 6 and Table S4, Chatterjee's correlation coefficients show that precipitation has the strongest relationship with forest GPP in the Qinghai-Tibet Plateau, with a correlation coefficient of 0.44. This makes it one of the top three factors that affect forest GPP in semi-arid, temperate semi-humid, and arid regions. This strong correlation highlights the importance of water availability in areas where vegetation is highly sensitive to water stress. In the Qinghai-Tibet Plateau and arid and semi-arid regions, precipitation levels are generally low, and vegetation in these areas frequently experiences drought stress. Drought limits photosynthesis in leaves through stomatal closure and metabolic impairment (Lin et al., 2024). Therefore, precipitation plays a critical role in determining forest GPP in these regions. In cold semi-humid, and humid regions, precipitation ranks fourth, indicating that it remains one of the primary limiting factors for forest GPP in these climate zones. However, other factors, such as temperature and radiation, might play a more prominent role in these regions.

Temperature is one of the three primary factors exhibiting a robust correlation with forest GPP throughout all climate zones, except for the Qinghai-Tibet Plateau and arid regions. In particular, in humid zones, temperature ranks as the most important determinant of GPP, with a correlation coefficient of 0.162. The abundant water supply in these regions alleviates the constraints of water stress, making temperature the dominant limiting factor for plant growth and photosynthetic efficiency. Additionally, temperature interacts with other environmental variables, such as solar radiation, further enhancing GPP. For instance, in humid zones, solar radiation is the second most important factor, while in cold semi-humid regions, solar radiation is the most important determinant, with a correlation coefficient of 0.183.

DEM is one of the three most important factors that determine forest GPP in the Qinghai-Tibet Plateau, semi-arid, and humid regions. Elevation-induced differences in temperature, rainfall, and solar radiation significantly influence these areas. For example, in the Qinghai-Tibet Plateau, solar radiation associated with DEM is the second most important factor. This is because topographic features are considered critical in shaping local microclimates. Soil moisture exhibits the third-largest association with forest GPP in the cold semi-humid zone and the fourth highest in the Qinghai-Tibet Plateau, semi-arid, and arid regions. In arid regions, the water-related factor of soil moisture, in addition to precipitation, plays a critical role in determining GPP. Soil moisture is the most direct and primary source of water for vegetation, and its dynamic changes significantly influence vegetation biomass and phenology (Luo et al., 2021). FFI ranks among the top three contributors to forest GPP in arid and temperate semi-humid regions. Forest fragmentation enhances edge effects, exposing more forest boundaries to external environments, resulting in significant changes. Microclimatic changes at the forest edge, such as decreased humidity and increased light exposure, primarily drive these effects (Dantas de Paula et al., 2015). A substantial correlation exists between nocturnal illumination and forest GPP across all climatic zones, reflecting the influence of urbanization and land-use changes on forest productivity. The correlation between slope and forest GPP is weak and statistically insignificant on the Qinghai-Tibet Plateau and in semi-arid and temperate semi-humid regions. All climate zones show the weakest relationships with aspect, and there is no significant connection between this variable and forest GPP in the Qinghai-Tibet Plateau, semi-arid, and temperate semi-humid

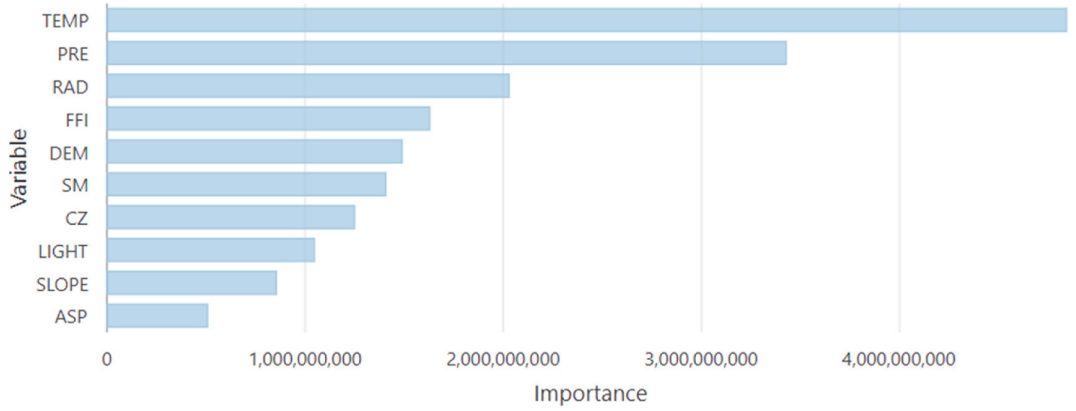


Fig. 5. Summary of variable importance to explain forest GPP in China (1990–2018) according to RF mean decrease impurity values (TEMP: Mean annual temperature, PRE: Mean annual precipitation, RAD: Solar radiation, FFI: Forest Fragmentation Index, DEM: Elevation, SM: Soil moisture, CZ: Climate zone, LIGHT: Nighttime lights, SLOPE: Slope, ASP: Aspect).

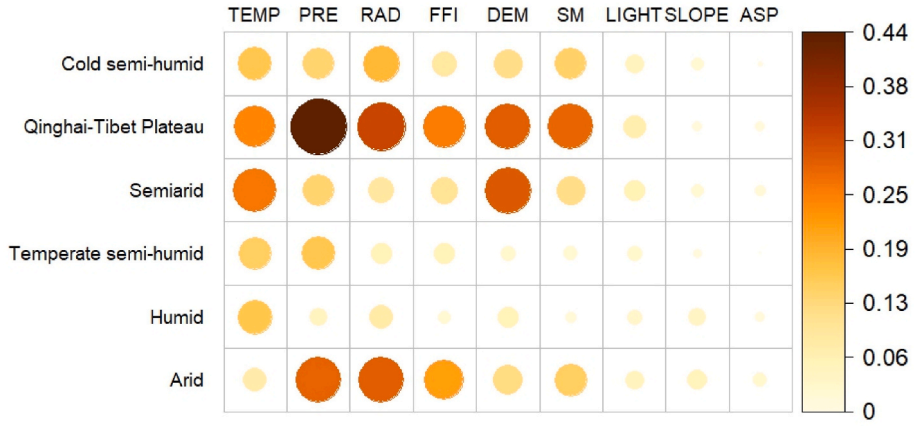


Fig. 6. Chatterjee's correlation coefficient between each driver (TEMP: Mean annual temperature, PRE: Mean annual precipitation, RAD: Solar radiation, FFI: Forest Fragmentation Index, DEM: Elevation, SM: Soil moisture, LIGHT: Nighttime lights, SLOPE: Slope, ASP: Aspect) and forest GPP in each climate zone.

regions.

4. Discussion

4.1. Quantification of forest GPP drivers in China and its climate zones

This work improves the existing knowledge about forest GPP in China and its climate zones from 1990 to 2018. Besides using new variables in the machine learning modeling process to better account for climate and anthropogenic effects, it provides the driver's signature of each climate zone using a novel correlation coefficient.

Results show that the mean annual forest GPP was 914.74 gC m^{-2} y^{-1} during the study period in China. We also found that forest GPP showed a significantly increasing trend ($p < 0.01$) at a rate of 4.09 gC m^{-2} y^{-1}. Our study solely concentrates on forested areas, making it impossible to directly compare these values with those from other studies that also include non-forested areas in China. Using light use efficiency, process-based, and machine learning methods, GPP has been estimated for China from 300 to 1277 gC m^{-2} y^{-1} for different time periods based on more than 20 studies (Li et al., 2016). Using three decades of time-series modeling based on the optimum light use efficiency and autoregressive integrated moving average (ARIMA) model, Bo et al. (2022) found that GPP increased 2.268 gC m^{-2} y^{-1} in China. Additionally, Ma et al. (2019) estimated a decline in GPP increment rate for China from 18.82 gC m^{-2} y^{-1} in 2000–2008 to 3.48 gC m^{-2} y^{-1} in 2008–2016. When looking at GPP studies from other countries, such as India, high increasing rates between 5.36 and 6.44 gC m^{-2} y-1 have also been reported (Gupta et al., 2023). Since our study did not include the non-forested areas, it is expected that these values are lower than the ones found in our study since we did not account for the negative impact of urbanization in China. These differences may also be justified by the different datasets and modeling approaches used, highlighting substantial uncertainties in literature. The results presented in this study disambiguate these uncertainties by separating the forested from non-forested areas, thereby focusing solely on the forested areas of China.

China denotes a southeast-northwest downward trend in average forest GPP during the study period, which is consistent with previous studies on GPP (Li et al., 2016). There are persistent cold spots in areas with higher altitudes and latitudes, such as in the Qinghai-Tibet Plateau. The low temperature affects the activity of photosynthesis-related enzymes in vegetation, thereby limiting the photosynthetic rate (Keenan and Riley, 2018). At the same time, solar radiation has an inhibitory effect on vegetation in the Qinghai-Tibet Plateau, probably due to solar radiation, which causes the melting of snow and permafrost, increases the water content of vegetation roots, and puts the soil in an anaerobic state, thereby reducing vegetation productivity (Yang et al., 2015). As the world warms, the demand for water in arid vegetation increases significantly, making the GPP in these areas more susceptible to strong changes in precipitation (Lai et al., 2022). This may explain the diminishing cold spots and intensifying cold spots that appear in the arid and semi-arid regions in places such as Inner Mongolia and the Northwest. The humid region, the southern and eastern parts of the semi-humid region, and the western and northeastern parts of the cold semi-humid region are concentrated areas of hotspots, and the hotspots of forest GPP continue to intensify. In areas with sufficient precipitation and moderate temperatures during the growing season, forest GPP will increase significantly. These regions may experience greater forest GPP changes due to changes in precipitation conditions, so sporadic hotspots are usually distributed around the edges of persistent hotspots and enhanced hotspots. At the same time, the positive contributions of human activities such as the Grain for Green Program launched by the Chinese government in 1999 with the intention of converting steep-sloped and degraded agriculture and barren land into forest to mitigate soil erosion, enhance biodiversity, and conserve natural resources (Song et al., 2014) have also significantly increased the GPP in these areas. The central part of the semi-arid and semi-humid region, such as the Loess Plateau, has seen the formation of diminishing cold spots. This transformation of farmland into forests or grasslands has improved carbon sequestration and ecosystem productivity (Lu et al., 2023). Planners can effectively use the spatiotemporal trends of forest GPP to prioritize mitigation policies, allocating more resources to areas they deem more important for application.

As with other studies that looked at GPP, the RF model did better than the XGBoost in terms of model validation (Chang et al., 2023). The most important determinants of forest GPP identified in this study, except for FFI and climate zones that were never used before, confirm the temperature as being the most important driver of GPP. Yao et al. (2018) found that the dominant driver of GPP is temperature, followed by precipitation and solar radiation. Chang et al. (2023) reported that temperature and vegetation indices were the most important variables, followed by solar radiation and moisture variables. In our study, while mean temperature, the most important driver (26.5 %), had a significant increasing rate during the study period (0.019 °C), precipitation, the second most important driver (18.6 %), exhibited only a slight upward trend (0.936 mm), which was not statistically significant. On the other hand, solar radiation, the third most important driver (11 %), decreased by 0.006 W m^{-2} and was also not statistically significant. This suggests that the increase in the mean temperature in China has been driving the changes in forest GPP during the study period. Machine learning approaches are confirmed to be a straightforward way of modeling spatiotemporal complex phenomena, such as forest GPP or specific types

of forests like mangroves (Sun et al., 2024), that can be easily adapted to include additional variables and/or to extend the temporal period of analysis.

This study further analyzed forest GPP by climate zones, utilizing the novel Chatterjee's correlation (Chatterjee, 2021) to identify the strength of the association of each driver with forest GPP. Each climate zone exhibited unique driver signatures in terms of their order and importance. Thus, the climate zones have different characteristics that affect forest GPP. This means that the study of drivers should take this difference in climate zones into account.

Interestingly, the FFI emerged as an important driver in explaining forest GPP in China. While forest GPP increased, FFI decreased during the same period studied. Given that human activities primarily cause forest disturbances (Haddad et al., 2015), a reduction in FFI could indicate an increase in plant photosynthesis efficiency. At the climate zone level, Chatterjee's correlations showed that FFI was the third most dominant driver in the arid and temperate semi-humid climate zones, respectively. Therefore, forest management may play a crucial role in determining forest GPP in these climate zones, particularly in the arid zone where numerous spatiotemporal cold spots of forest GPP exist.

4.2. Limitations and future research

The insights provided by this study, although valuable, need to consider some limitations inherent to uncertainties related to the use of temporal series of remote sensing data as well as data processing and modeling. Despite the existence of studies using the non-parametric Mann-Kendall test to study time series data (Wang et al., 2023), this method does not identify non-linear trends (Hirschand Slack, 1984) that are common in environmental series, such as the ones used in this study.

We only used a subset of the available variables, taking into account a literature review analysis and the availability of data. Other variables, such as detailed types of forests, or the same variables using a different spatial resolution, would impact the modeling results. It is crucial to emphasize that machine learning approaches offer several adjustable parameters, including the number of trees, leaf sizes, and the percentage of the validation sample, among others. After conducting numerous experiments, including parameter optimization, we retained the parameters that best matched the results of the validation dataset. However, the inclusion of additional variables or datasets with different characteristics, including seasonal dynamics effects (Wang et al., 2020), would imply finding new optimal parameters, preventing the direct replicability of these parameters in other studies.

Another important aspect to consider when using machine learning is the difficulty in interpreting simulation results, especially when trying to analyze the data trends and explain the potential connections to climate change processes and ecosystems (Wu et al., 2019). Despite delivering faster outcomes than alternative methods, solely relying on machine learning models fails to adequately elucidate the physiological mechanisms associated with vegetative processes (Liao et al., 2023). To minimize this issue, this study included trend analysis and spatiotemporal statistical methods. However, further studies are required to better understand the intricate relationship between forest GPP and explanatory variables.

Increasingly, environmental, climate, topographic, and anthropogenic data are being made easier to access by users at no cost. At the same time, new algorithms and data-based methodological approaches open new avenues to explore larger amounts of datasets, providing new perspectives on spatiotemporal phenomena such as GPP. Despite the existing limitations, this study adds a new perspective based on the use of different and recent spatiotemporal methods and variables. Future research could include the use of additional methods, such as SHAP interaction values (Lundberg et al., 2020), and/or the use of Automated Machine Learning (AML) (Gaber et al., 2024) to estimate past or future scenarios of GPP productivity (Uchale et al., 2023).

5. Conclusion

To conclude, this study quantified the importance of drivers of forest GPP in China and its climate zones from 1990 to 2018. Making use of long-term temporal series of open data based on remote sensing, we started by analyzing the trends of forest GPP and its drivers, including forest fragmentation and climate zones. Then, using a random forest model, we estimated the most important drivers. Results showed that the mean annual forest GPP was 914.74 gC m^{-2} y^{-1} and had an increasing trend at a rate of 4.09 gC m^{-2} y^{-1}. The mean temperature was the most important driver (26.5 %), followed by precipitation (18.6 %) and solar radiation (11 %). Both estimated forest GPP values and most important drivers concur with the results of other studies that modeled GPP despite using different modeling approaches and drivers. Further statistical and spatiotemporal analysis showed that each climate zone has a unique signature regarding the order and importance of drivers, which may provide important insights to decision-makers to define more effective management strategies. This study also highlights the importance of forest fragmentation, the third most important driver in the temperate semi-humid and arid zones, suggesting that special attention should be given to sustainable forest management in those areas.

CRediT authorship contribution statement

Chenxi Zhu: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Guojie Wang:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization. **Yuhao Shao:** Writing – review & editing, Data curation. **Wen Dai:** Writing – review & editing, Resources. **Qiang Liu:** Writing – review & editing, Validation, Methodology. **Shuangao Wang:** Writing – review & editing, Resources. **Ana Cristina Costa:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Pedro Cabral:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded by the National Natural Science Foundation of China (#42275028). The article was partially supported by national funds through FCT (Fundação para a Ciência e a Tecnologia), under the project - UIDB/04152/2020 (DOI: 10.54499/UIDB/04152/2020) - Centro de Investigação em Gestão de Informação (MagIC)/NOVA IMS).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclepro.2025.145616.

Data availability

The dataset with processed data used to model forest GPP is available in:https://doi.org/10.6084/m9.figshare.28830437.v1.

References

Anderson, T.W., Darling, D.A., 1954. A test of goodness of fit. J. Am. Stat. Assoc. 49, 765–769. https://doi.org/10.1080/01621459.1954.10501232.

Beaver, J., Humphreys, E.R., King, D., 2024. Random forest development and modeling of gross primary productivity in the Hudson Bay lowlands. Can. J. Rem. Sens. 50. https://doi.org/10.1080/07038992.2024.2355937.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M.A., Baldocchi, D., Bonan, G.B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K.W., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F.I., Papale, D., 2010. Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. Science 329, 834–838. https://doi.org/10.1126/science.1184984, 1979.

Bentéjac, C., Csörgő, A., Martínez-Muñoz, G., 2021. A comparative analysis of gradient boosting algorithms. Artif. Intell. Rev. 54, 1937–1967. https://doi.org/10.1007/s10462-020-09896-5.

Bo, Y., Li, X., Liu, K., Wang, S., Zhang, H., Gao, X., Zhang, X., 2022. Three decades of gross primary production (GPP) in China: variations, trends, attributions, and prediction inferred from multiple datasets and time series modeling. Remote Sens. 14, 2564. https://doi.org/10.3390/rs14112564.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. Classification and Regression Trees. Routledge. https://doi.org/10.1201/9781315139470.

CGIAR-CSI, 2022. SRTM DEM dataset in China (2000). National Tibetan Plateau Data Center.

Chang, X., Xing, Y., Gong, W., Yang, C., Guo, Z., Wang, D., Wang, J., Yang, H., Xue, G., Yang, S., 2023. Evaluating gross primary productivity over 9 ChinaFlux sites based on random forest regression models, remote sensing, and eddy covariance data. Sci. Total Environ. 875, 162601. https://doi.org/10.1016/j.scitotenv.2023.162601.

Chatterjee, S., 2021. A new coefficient of correlation. J. Am. Stat. Assoc. 116, 2009–2022. https://doi.org/10.1080/01621459.2020.1758115.

Chen, S., Ma, M., Wu, S., Tang, Q., Wen, Z., 2023. Topography intensifies variations in the effect of human activities on forest NPP across altitude and slope gradients. Environ. Dev. 45, 100826. https://doi.org/10.1016/j.envdev.2023.100826.

Chen, T., Guestrin, C., 2016. XGBoost. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Chen, T., Meunier, F., Peaucelle, M., Tang, G., Yuan, Y., Verbeeck, H., 2024. Elevated atmospheric CO_2 concentration and vegetation structural changes contributed to gross primary productivity increase more than climate and forest cover changes in subtropical forests of China. Biogeosciences 21, 2253–2272. https://doi.org/10.5194/bg-21-2253-2024.

Dantas de Paula, M., Groeneveld, J., Huth, A., 2015. Tropical forest degradation and recovery in fragmented landscapes — simulating changes in tree community, forest hydrology and carbon balance. Glob Ecol Conserv 3, 664–677. https://doi.org/10.1016/j.gecco.2015.03.004.

Deng, X., Gao, F., Liao, S., Liu, Y., Chen, W., 2023. Spatiotemporal evolution patterns of urban heat island and its relationship with urbanization in Guangdong-Hong Kong-Macao greater Bay area of China from 2000 to 2020. Ecol. Indic. 146, 109817. https://doi.org/10.1016/j.ecolind.2022.109817.

Ding, L., Li, Z., Wang, X., Shen, B., Xiao, L., Dong, G., Yu, L., Nandintsetseg, B., Shi, Z., Chang, J., Shao, C., 2024. Spatiotemporal patterns and driving factors of gross primary productivity over the Mongolian Plateau steppe in the past 20 years. Sci. Total Environ. 920, 170886. https://doi.org/10.1016/j.scitotenv.2024.170886.

Dwass, M., 1960. Some k-sample rank-order tests. In: Olkin, I., Ghurye, S.G., Hoeffding, H., Madow, W.G., Mann, H.B. (Eds.), Contributions to Probability and Statistics. Stanford University Press, pp. 198–202.

ESRI, 2023. ArcGIS Pro 3.3.

Gaber, M., Kang, Y., Schurgers, G., Keenan, T., 2024. Using automated machine learning for the upscaling of gross primary productivity. Biogeosciences 21, 2447–2472. https://doi.org/10.5194/bg-21-2447-2024.

Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. Geogr. Anal. 24, 189–206. https://doi.org/10.1111/j.1538-4632.1992.tb00261.x.

Guo, Heyi, Cao, C., Xu, M., Yang, X., Chen, Y., Wang, K., Duerler, R.S., Li, J., Gao, X., 2023. Spatiotemporal distribution pattern and driving factors analysis of GPP in Beijing-Tianjin-Hebei region by long-term MODIS data. Remote Sens. 15, 622. https://doi.org/10.3390/rs15030622.

Guo, Hui, Zhou, X., Dong, Y., Wang, Y., Li, S., 2023. On the use of machine learning methods to improve the estimation of gross primary productivity of maize field with drip irrigation. Ecol. Model. 476, 110250. https://doi.org/10.1016/j.ecolmodel.2022.110250.

Gupta, S., Deb Burman, P.K., Tiwari, Y.K., Dumka, U.C., Kumari, N., Srivastava, A., Raghubanshi, A.S., 2023. Understanding carbon sequestration trends using model and satellite data under different ecosystems in India. Sci. Total Environ. 897, 166381. https://doi.org/10.1016/j.scitotenv.2023.166381.

Haddad, N.M., Brudvig, L.A., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D., Lovejoy, T.E., Sexton, J.O., Austin, M.P., Collins, C.D., Cook, W.M., Damschen, E.I., Ewers, R.M., Foster, B.L., Jenkins, C.N., King, A.J., Laurance, W.F., Levey, D.J., Margules, C.R., Melbourne, B.A., Nicholls, A.O., Orrock, J.L., Song, D.-X., Townshend, J.R., 2015. Habitat fragmentation and its lasting impact on Earth's ecosystems. Sci. Adv. 1. https://doi.org/10.1126/sciadv.1500052.

Hao, A., Duan, H., Wang, X., Zhao, G., You, Q., Peng, F., Du, H., Liu, F., Li, C., Lai, C., Xue, X., 2021. Different response of alpine meadow and alpine steppe to climatic and anthropogenic disturbance on the Qinghai-Tibetan Plateau. Glob Ecol Conserv 27, e01512. https://doi.org/10.1016/j.gecco.2021.e01512.

He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., Li, X., 2020. The first high-resolution meteorological forcing dataset for land process studies over China. Sci. Data 7, 25. https://doi.org/10.1038/s41597-020-0369-y.

He, J., Zhang, W., Yang, M., 2024. The spatial and temporal characteristics of urban public safety under the residents' complaints: evidence from 12345 data in Beijing, China. Journal of Urban Management 13, 217–231. https://doi.org/10.1016/j.jum.2024.01.003.

Hesselbarth, M.H.K., Sciaini, M., With, K.A., Wiegand, K., Nowosad, J., 2019. *Landscapemetrics* : an open-source *R* tool to calculate landscape metrics. Ecography 42, 1648–1657. https://doi.org/10.1111/ecog.04617.

Hirsch, R.M., Slack, J.R., 1984. A nonparametric trend test for seasonal data with serial dependence. Water Resour. Res. 20, 727–732. https://doi.org/10.1029/WR020i006p00727.

Hoover, D.L., Lauenroth, W.K., Milchunas, D.G., Porensky, L.M., Augustine, D.J., Derner, J.D., 2021. Sensitivity of productivity to precipitation amount and pattern varies by topographic position in a semiarid grassland. Ecosphere 12. https://doi.org/10.1002/ecs2.3376.

Hu, X., Shi, L., Lin, L., Li, S., Deng, X., Li, L., Bian, J., Lian, X., 2024. A novel hybrid modelling framework for GPP estimation: integrating a multispectral surface reflectance based V_{cmax}^{25} simulator into the process-based model. Sci. Total Environ. 921, 171182. https://doi.org/10.1016/j.scitotenv.2024.171182.

Hudak, A.T., Fekety, P.A., Kane, V.R., Kennedy, R.E., Filippelli, S.K., Falkowski, M.J., Tinkham, W.T., Smith, A.M.S., Crookston, N.L., Domke, G.M., Corrao, M.V., Bright, B.C., Churchill, D.J., Gould, P.J., McGaughey, R.J., Kane, J.T., Dong, J., 2020. A carbon monitoring system for mapping regional, annual aboveground biomass across the northwestern USA. Environ. Res. Lett. 15, 095003. https://doi.org/10.1088/1748-9326/ab93f9.

Hunter, F.D.L., Mitchard, E.T.A., Tyrrell, P., Russell, S., 2020. Inter-seasonal time series imagery enhances classification accuracy of grazing resource and land degradation maps in a savanna ecosystem. Remote Sens. 12, 198. https://doi.org/10.3390/rs12010198.

Ishwaran, H., 2015. The effect of splitting on random forests. Mach. Learn. 99, 75–118. https://doi.org/10.1007/s10994-014-5451-2.

Jemeljanova, M., Kmoch, A., Uuemaa, E., 2024. Adapting machine learning for environmental spatial data - a review. Ecol. Inform. 81, 102634. https://doi.org/10.1016/j.ecoinf.2024.102634.

Jung, M., Vetter, M., Herold, M., Churkina, G., Reichstein, M., Zaehle, S., Ciais, P., Viovy, N., Bondeau, A., Chen, Y., Trusilova, K., Feser, F., Heimann, M., 2007. Uncertainties of modeling gross primary productivity over Europe: a systematic study on the effects of using different drivers and terrestrial biosphere models. Glob. Biogeochem. Cycles 21. https://doi.org/10.1029/2006GB002915.

Keenan, T.F., Riley, W.J., 2018. Greening of the land surface in the world's cold regions consistent with recent warming. Nat. Clim. Change 8, 825–828. https://doi.org/10.1038/s41558-018-0258-y.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World map of the köppen-geiger climate classification updated. Meteorol. Z. 15, 259–263. https://doi.org/10.1127/0941-2948/2006/0130.

Kruskal, W.H., Wallis, W.A., 1952. Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. 47, 583–621. https://doi.org/10.1080/01621459.1952.10483441.

Lai, W., Wang, M., Wei, J., Zhang, J., Song, J., Zhou, H., Chou, S., Wang, Y., 2022. Separating the impact of climate changes and human activities on vegetation growth based on the NDVI in China. Adv. Meteorol. 2022, 1–11. https://doi.org/10.1155/2022/6294029.

Li, X., Zhu, Z., Zeng, H., Piao, S., 2016. Estimation of gross primary production in China (1982–2010) with multiple ecosystem models. Ecol. Model. 324, 33–44. https://doi.org/10.1016/j.ecolmodel.2015.12.019.

Liao, D., Niu, J., Ciais, P., Du, T., Zhang, B., Kang, S., 2024. Changing climate threatens irrigation benefits of maize gross primary productivity in China. Earths Future 12. https://doi.org/10.1029/2022EF003474.

Liao, Z., Zhou, B., Zhu, J., Jia, H., Fei, X., 2023. A critical review of methods, principles and progress for estimating the gross primary productivity of terrestrial ecosystems. Front. Environ. Sci. 11. https://doi.org/10.3389/fenvs.2023.1093095.

Lin, Y., Xie, T., Li, S., Li, X., Liu, W., 2024. Amplified photosynthetic responses to drought events offset the positive effects of warming on arid desert plants. Sci. Total Environ. 952, 175899. https://doi.org/10.1016/j.scitotenv.2024.175899.

Lu, J., Wang, G., Feng, D., Nooni, I.K., 2023. Improving the gross primary production estimate by merging and downscaling based on deep learning. Forests 14, 1201. https://doi.org/10.3390/f14061201.

Lu, Q., Liu, H., Wei, L., Zhong, Y., Zhou, Z., 2024. Global prediction of gross primary productivity under future climate change. Sci. Total Environ. 912, 169239. https://doi.org/10.1016/j.scitotenv.2023.169239.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9.

Luo, M., Meng, F., Sa, C., Duan, Y., Bao, Y., Liu, T., De Maeyer, P., 2021. Response of vegetation phenology to soil moisture dynamics in the mongolian Plateau. Catena 206, 105505. https://doi.org/10.1016/j.catena.2021.105505.

Ma, J., Li, J., Wu, W., Liu, J., 2023. Global forest fragmentation change from 2000 to 2020. Nat. Commun. 14, 3752. https://doi.org/10.1038/s41467-023-39221-x.

Ma, J., Xiao, X., Miao, R., Li, Y., Chen, B., Zhang, Y., Zhao, B., 2019. Trends and controls of terrestrial gross primary productivity of China during 2000–2016. Environ. Res. Lett. 14, 084032. https://doi.org/10.1088/1748-9326/ab31e4.

Mann, H.B., 1945. Nonparametric tests against trend. Econometrica 13, 245. https://doi.org/10.2307/1907187.

Meyer, H., Pebesma, E., 2022. Machine learning-based global maps of ecological variables and the challenge of assessing them. Nat. Commun. 13, 2208. https://doi.org/10.1038/s41467-022-29838-9.
Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 37, 17. https://doi.org/10.2307/2332142.
Moran, P.A.P., 1948. The interpretation of statistical maps. J R Stat Soc Series B Stat Methodol 10, 243–251. https://doi.org/10.1111/j.2517-6161.1948.tb00012.x.
Morreale, L.L., Thompson, J.R., Tang, X., Reinmann, A.B., Hutyra, L.R., 2021. Elevated growth and biomass along temperate forest edges. Nat. Commun. 12, 7181. https://doi.org/10.1038/s41467-021-27373-7.
Palmer, J.G., Verdon-Kidd, D., Allen, K.J., Higgins, P., Cook, B.I., Cook, E.R., Turney, C.S. M., Baker, P.J., 2024. Drought and deluge: the recurrence of hydroclimate extremes during the past 600 years in eastern Australia's natural resource management (NRM) clusters. Nat. Hazards 120, 3565–3587. https://doi.org/10.1007/s11069-023-06288-0.
Pan, Y., Yang, R., Qiu, J., Wang, J., Wu, J., 2023. Forty-year spatio-temporal dynamics of agricultural climate suitability in China reveal shifted major crop production areas. Catena 226, 107073. https://doi.org/10.1016/j.catena.2023.107073.
Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Reichstein, M., Ribeca, A., van Ingen, C., Vuichard, N., Zhang, L., Amiro, B., Ammann, C., Arain, M.A., Ardö, J., Arkebauer, T., Arndt, S.K., Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L.B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T.A., Blanken, P.D., Bohrer, G., Boike, J., Bolstad, P.V., Bonal, D., Bonnefond, J.-M., Bowling, D.R., Bracho, R., Brodeur, J., Brümmer, C., Buchmann, N., Burban, B., Burns, S.P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T.R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B.D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P.S., D'Andrea, E., da Rocha, H., Dai, X., Davis, K.J., Cinti, B. De, Grandcourt, A. de, Ligne, A. De, De Oliveira, R.C., Delpierre, N., Desai, A.R., Di Bella, C.M., Tommasi, P. di, Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrêne, E., Dunn, A., Dušek, J., Eamus, D., Eichelmann, U., ElKhidir, H.A.M., Eugster, W., Ewenz, C.M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., Gharun, M., Gianelle, D., Gielen, B., Gioli, B., Gitelson, A., Goded, I., Goeckede, M., Goldstein, A.H., Gough, C.M., Goulden, M.L., Graf, A., Griebel, A., Gruening, C., Grünwald, T., Hammerle, A., Han, S., Han, X., Hansen, B.U., Hanson, C., Hatakka, J., He, Y., Hehn, M., Heinesch, B., Hinko-Najera, N., Hörtnagl, L., Hutley, L., Ibrom, A., Ikawa, H., Jackowicz-Korczynski, M., Janouš, D., Jans, W., Jassal, R., Jiang, S., Kato, T., Khomik, M., Klatt, J., Knohl, A., Knox, S., Kobayashi, H., Koerber, G., Kolle, O., Kosugi, Y., Kotani, A., Kowalski, A., Kruijt, B., Kurbatova, J., Kutsch, W.L., Kwon, H., Launiainen, S., Laurila, T., Law, B., Leuning, R., Li, Yingnian, Liddell, M., Limousin, J.-M., Lion, M., Liska, A.J., Lohila, A., López-Ballesteros, A., López-Blanco, E., Loubet, B., Loustau, D., Lucas-Moffat, A., Lüers, J., Ma, S., Macfarlane, C., Magliulo, V., Maier, R., Mammarella, I., Manca, G., Marcolla, B., Margolis, H.A., Marras, S., Massman, W., Mastepanov, M., Matamala, R., Matthes, J.H., Mazzenga, F., McCaughey, H., McHugh, I., McMillan, A.M.S., Merbold, L., Meyer, W., Meyers, T., Miller, S.D., Minerbi, S., Moderow, U., Monson, R.K., Montagnani, L., Moore, C.E., Moors, E., Moreaux, V., Moureaux, C., Munger, J.W., Nakai, T., Neirynck, J., Nesic, Z., Nicolini, G., Noormets, A., Northwood, M., Nosetto, M., Nouvellon, Y., Novick, K., Oechel, W., Olesen, J.E., Ourcival, J.-M., Papuga, S.A., Parmentier, F.-J., Paul-Limoges, E., Pavelka, M., Peichl, M., Pendall, E., Phillips, R.P., Pilegaard, K., Pirk, N., Posse, G., Powell, T., Prasse, H., Prober, S.M., Rambal, S., Rannik, Ü., Raz-Yaseef, N., Rebmann, C., Reed, D., Dios, V.R. de, Restrepo-Coupe, N., Reverter, B.R., Roland, M., Sabbatini, S., Sachs, T., Saleska, S.R., Sánchez-Cañete, E.P., Sanchez-Mejia, Z.M., Schmid, H.P., Schmidt, M., Schneider, K., Schrader, F., Schroder, I., Scott, R.L., Sedlák, P., Serrano-Ortíz, P., Shao, C., Shi, P., Shironya, I., Siebicke, L., Šigut, L., Silberstein, R., Sirca, C., Spano, D., Steinbrecher, R., Stevens, R.M., Sturtevant, C., Suyker, A., Tagesson, T., Takanashi, S., Tang, Y., Tapper, N., Thom, J., Tomassucci, M., Tuovinen, J.-P., Urbanski, S., Valentini, R., van der Molen, M., van Gorsel, E., van Huissteden, K., Varlagin, A., Verfaillie, J., Vesala, T., Vincke, C., Vitale, D., Vygodskaya, N., Walker, J.P., Walter-Shea, E., Wang, H., Weber, R., Westermann, S., Wille, C., Wofsy, S., Wohlfahrt, G., Wolf, S., Woodgate, W., Li, Yuelin, Zampedri, R., Zhang, J., Zhou, G., Zona, D., Agarwal, D., Biraud, S., Torn, M., Papale, D., 2020. The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Sci. Data 7, 225. https://doi.org/10.1038/s41597-020-0534-3.
Peng, S., Ding, Y., Liu, W., Li, Z., 2019. 1 km monthly temperature and precipitation dataset for China from 1901 to 2017. Earth Syst. Sci. Data 11, 1931–1946. https://doi.org/10.5194/essd-11-1931-2019.
Prakash Sarkar, D., Uma Shankar, B., Ranjan Parida, B., 2022. Machine learning approach to predict terrestrial gross primary productivity using topographical and remote sensing data. Ecol. Inform. 70, 101697. https://doi.org/10.1016/j.ecoinf.2022.101697.
Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1.
Sen, P.K., 1968. Estimates of the regression coefficient based on kendall's tau. J. Am. Stat. Assoc. 63, 1379–1389. https://doi.org/10.1080/01621459.1968.10480934.
Song, X., Peng, C., Zhou, G., Jiang, H., Wang, W., 2014. Chinese grain for green program led to highly increased soil organic carbon levels: a meta-analysis. Sci. Rep. 4, 4460. https://doi.org/10.1038/srep04460.
Steel, R.G.D., 1961. Some rank sum multiple comparisons tests. Biometrics 17, 539. https://doi.org/10.2307/2527854.
Steel, R.G.D., 1960. A rank sum test for comparing all pairs of treatments. Technometrics 2, 197–207. https://doi.org/10.1080/00401706.1960.10489894.
Sun, Z., An, Y., Kong, J., Zhao, J., Cui, W., Nie, T., Zhang, T., Liu, W., Wu, L., 2024. Exploring the spatio-temporal patterns of global mangrove gross primary production and quantifying the factors affecting its estimation, 1996–2020. Sci. Total Environ. 908, 168262. https://doi.org/10.1016/j.scitotenv.2023.168262.
Theil, H., 1992. A rank-invariant method of linear and polynomial regression analysis, 345–381. https://doi.org/10.1007/978-94-011-2546-8_20.
Theil, H., 1950. A rank-invariant method of linear and polynomial regression analysis, 3; confidence regions for the parameters of polynomial regression equations. Indagat. Math. 1.
Uchale, G., Deb Burman, P.K., Tiwari, Y.K., Datye, A., Sarkar, A., 2023. Investigating terrestrial carbon uptake over India using multimodel simulations of gross primary productivity and satellite-based biophysical product. J Geophys Res Biogeosci 128. https://doi.org/10.1029/2023JG007468.
Wang, B., Yue, X., Zhou, H., Zhu, J., 2022. Impact of diffuse radiation on evapotranspiration and its coupling to carbon fluxes at global FLUXNET sites. Agric. For. Meteorol. 322, 109006. https://doi.org/10.1016/j.agrformet.2022.109006.
Wang, G., Peng, W., Zhang, L., Zhang, J., 2023. Quantifying the impacts of natural and human factors on changes in NPP using an optimal parameters-based geographical detector. Ecol. Indic. 155, 111018. https://doi.org/10.1016/j.ecolind.2023.111018.
Wang, L., Tian, F., Huang, K., Wang, Y., Wu, Z., Fensholt, R., 2020. Asymmetric patterns and temporal changes in phenology-based seasonal gross carbon uptake of global terrestrial ecosystems. Global Ecol. Biogeogr. 29, 1020–1033. https://doi.org/10.1111/geb.13084.
Wang, S., Zhang, Y., Ju, W., Qiu, B., Zhang, Z., 2021. Tracking the seasonal and inter-annual variations of global gross primary production during last four decades using satellite near-infrared reflectance data. Sci. Total Environ. 755, 142569. https://doi.org/10.1016/j.scitotenv.2020.142569.
Wang, X., Geng, X., Liu, B., Cai, D., Li, D., Xiao, F., Zhu, B., Hua, T., Lu, R., Liu, F., 2022. Desert ecosystems in China: past, present, and future. Earth Sci. Rev. 234, 104206. https://doi.org/10.1016/j.earscirev.2022.104206.
Welp, L.R., Keeling, R.F., Meijer, H.A.J., Bollenbacher, A.F., Piper, S.C., Yoshimura, K., Francey, R.J., Allison, C.E., Wahlen, M., 2011. Interannual variability in the oxygen isotopes of atmospheric CO2 driven by El niño. Nature 477, 579–582. https://doi.org/10.1038/nature10421.
Wu, W., Gong, C., Li, X., Guo, H., Zhang, L., 2019. An online deep convolutional model of gross primary productivity and net ecosystem exchange estimation for global forests. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 12, 5178–5188. https://doi.org/10.1109/JSTARS.2019.2954556.
Xiao-Ying, W., Chun-Yu, Z., Qing-Yu, J., 2013. Impacts of climate change on forest ecosystems in Northeast China. Adv. Clim. Change Res. 4, 230–241. https://doi.org/10.3724/SP.J.1248.2013.230.
Yang, J., Huang, X., 2021. The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019. Earth Syst. Sci. Data 13, 3907–3925. https://doi.org/10.5194/essd-13-3907-2021.
Yang, Y., Guan, H., Shen, M., Liang, W., Jiang, L., 2015. Changes in autumn vegetation dormancy onset date and the climate controls across temperate ecosystems in China from 1982 to 2010. Glob. Change Biol. 21, 652–665. https://doi.org/10.1111/gcb.12778.
Yang, Y., Nie, X., Cong, Z., Pan, X., Gong, X., Zhang, Z., 2024. Assessing the carbon-water compound use efficiency in fragile karst region: the Yunnan-Guizhou Plateau, China. Ecol. Indic. 166, 112320. https://doi.org/10.1016/j.ecolind.2024.112320.
Yao, Y., Wang, X., Li, Yue, Wang, T., Shen, M., Du, M., He, H., Li, Yingnian, Luo, W., Ma, M., Ma, Y., Tang, Y., Wang, H., Zhang, X., Zhang, Y., Zhao, L., Zhou, G., Piao, S., 2018. Spatiotemporal pattern of gross primary productivity and its covariation with climate in China over the last thirty years. Glob. Change Biol. 24, 184–196. https://doi.org/10.1111/gcb.13830.
Yap, B.W., Sim, C.H., 2011. Comparisons of various types of normality tests. J. Stat. Comput. Simulat. 81, 2141–2155. https://doi.org/10.1080/00949655.2010.520163.
Yin, C., Luo, M., Meng, F., Sa, C., Yuan, Z., Bao, Y., 2022. Contributions of climatic and anthropogenic drivers to net primary productivity of vegetation in the Mongolian Plateau. Remote Sens. 14, 3383. https://doi.org/10.3390/rs14143383.
Yin, Y., Tang, Q., Wang, L., Liu, X., 2016. Risk and contributing factors of ecosystem shifts over naturally vegetated land under climate change in China. Sci. Rep. 6, 20905. https://doi.org/10.1038/srep20905.
Yu, P., Zhou, T., Luo, H., Liu, X., Shi, P., Zhao, X., Xiao, Z., Zhang, Y., Zhou, P., 2022. Interannual variation of gross primary production detected from optimal convolutional neural network at multi-timescale water stress. Remote Sens Ecol Conserv 8, 409–425. https://doi.org/10.1002/rse2.252.
Yujie, L., Hong, J., Xuanguang, L., 2023. Simulation of vegetation gross primary productivity and its application in Fujian province of China using remote sensing and CatBoost algorithm. Journal of Geo-information Science 25, 1908–1922. https://doi.org/10.12082/DQXXKX.2023.220623.
Zhang, K., Chen, H., Ma, N., Shang, S., Wang, Y., Xu, Q., Zhu, G., 2024. A global dataset of terrestrial evapotranspiration and soil moisture dynamics from 1982 to 2020. Sci. Data 11, 445. https://doi.org/10.1038/s41597-024-03271-7.
Zhang, L., Ren, Z., Chen, B., Gong, P., Xu, B., Fu, H., 2024. A prolonged artificial nighttime-light dataset of China (1984-2020). Sci. Data 11, 414. https://doi.org/10.1038/s41597-024-03223-1.

Zhang, X., Bi, J., Zhu, D., Meng, Z., 2024. Seasonal variation of net ecosystem carbon exchange and gross primary production over a Loess Plateau semi-arid grassland of northwest China. Sci. Rep. 14, 2916. https://doi.org/10.1038/s41598-024-52559-6.

Zhang, X., Zhou, Y., He, W., Ju, W., Liu, Y., Bi, W., Cheng, N., Wei, X., 2022. Land cover change instead of solar radiation change dominates the forest GPP increase during the recent phase of the Shelterbelt program for Pearl river. Ecol. Indic. 136, 108664. https://doi.org/10.1016/j.ecolind.2022.108664.

Zhu, X.-J., Yu, G.-R., Chen, Z., Zhang, W.-K., Han, L., Wang, Q.-F., Chen, S.-P., Liu, S.-M., Wang, H.-M., Yan, J.-H., Tan, J.-L., Zhang, F.-W., Zhao, F.-H., Li, Y.-N., Zhang, Y.-P., Shi, P.-L., Zhu, J.-J., Wu, J.-B., Zhao, Z.-H., Hao, Y.-B., Sha, L.-Q., Zhang, Y.-C., Jiang, S.-C., Gu, F.-X., Wu, Z.-X., Zhang, Y.-J., Zhou, L., Tang, Y.-K., Jia, B.-R., Li, Y.-Q., Song, Q.-H., Dong, G., Gao, Y.-H., Jiang, Z.-D., Sun, D., Wang, J.-L., He, Q.-H., Li, X.-H., Wang, F., Wei, W.-X., Deng, Z.-M., Hao, X.-X., Li, Y., Liu, X.-L., Zhang, X.-F., Zhu, Z.-L., 2023. Mapping Chinese annual gross primary productivity with eddy covariance measurements and machine learning. Sci. Total Environ. 857, 159390. https://doi.org/10.1016/j.scitotenv.2022.159390.