Evaluation of annual trends in carbon cycle variables simulated by CMIP6 Earth system models in China

Ziyang Li[1,2,3], Anzhou Zhao[3], Lidong Zou[1,2,*], Haigang Zhang[1], Feng Yue[4], Zhe Luo[2], Rui Bian[3], Ruihao Xu[3]

[1]Institute of Applied Artificial Intelligence of The Guangdong-Hongkong-Macao Greater Bay, Shenzhen Polytechnic University, Shenzhen 518055, China;
[2]School of Artificial Intelligence, Shenzhen Polytechnic University, Shenzhen 518055, China;
[3]School of Mining and Geomatics, Hebei University of Engineering, Handan 056038, China;
[4]School of Architecture and Urban Planning, Guangdong University of Technology, Guangzhou 510062, China

Correspondence to: Lidong Zou (lidongzou@szpu.edu.cn)

Abstract. Systematic evaluation of the carbon cycle physical and biological variables simulated in Earth System Model (ESM) participating in the Coupled Model Intercomparison Project Phase 6 (CMIP 6) is fundamental to the understanding of terrestrial ecosystems, as well as to future projections. Leaf Area Index (LAI), Gross Primary Productivity (GPP), Net Primary Productivity (NPP), Net Ecosystem Productivity (NEP) and Land Surface Temperature (LST) as key indicators of carbon cycle performance in ESM outputs, play a critical role in evaluating ecosystem functions. Assessing these metrics can provide valuable insights into the biases in model-simulated ecosystems and offer guidance for model improvement. In this study, we assessed the interannual trends performance of LAI, GPP, NPP, NEP and LST simulated by 12 CMIP6 ESMs during the historical period by using satellite LAI, NPP, NEP, LST and CSIF data as observations. The findings indicate that: (1) There are significant uncertainties in the overall trends in LAI, NPP, and LST captured by the CMIP6 ESM. Meanwhile, simulated GPP and NEP trends were lower than observations with discrepancies reaching $0.03 \cdot \text{yr}^{-1}$ for GPP and 2.46 g $C \cdot m^{-2} \cdot yr^{-1}$ for NEP. (2) Spatially, these models exhibited widespread misestimation of LAI, GPP, NPP, and NEP trends in most regions of China. The underestimation area except in limited areas, including parts of the western Tibetan Plateau, the Pearl River Basin in southern China, and the North China Plain. Meanwhile, the simulated LST trend is underestimated in northern China, while its overestimations in western and southern China. (3) ESMs inadequate responsiveness to anthropogenic and environmental forcing and incomplete mechanistic representation of plant respiration pathways struggled accurate simulation of trends in LAI, GPP, NPP, NEP and LST.

Key Words. CMIP6; Earth system model; Carbon cycle physical and biological variables; Vegetation Dynamic;

1 Introduction

Since the Industrial Revolution, human activities have significantly altered the structure of the atmosphere and terrestrial biosphere, leading to profound changes in the structure and functioning of terrestrial ecosystems (Allen et al., 2018). Vegetation, as a key component of the biosphere, regulates the exchange of carbon, water, momentum, and energy between

the land and the atmosphere (Wu et al., 2020). Vegetations maximize water-use efficiency at the leaf scale by dynamically regulating stomatal conductance, to effectively respond to rising atmospheric CO_2 concentrations and global warming (Fu et al., 2022). However, due to its high sensitivity to environmental conditions (Jung et al., 2017; Nemani et al., 2003), vegetation has undergone notable changes. The response of vegetation to global change has attracted significant attention in recent years (Chen et al., 2019; Hovenden et al., 2019; Zhu et al., 2016).

To capture the responses and feedbacks of terrestrial ecosystems to global environmental change, several Earth System Models (ESMs) in Coupled Model Intercomparison Project Phase 6 (CMIP6) simulate terrestrial vegetation structure and photosynthetic capacity (Song et al., 2021). Four key outputs of ESMs, which describe canopy structure and photosynthetic capacity—Leaf Area Index (LAI), Gross Primary Productivity (GPP), Net Primary Productivity (NPP), Net Ecosystem Productivity (NEP)—are critical components of the terrestrial carbon cycle (Le Quéré et al., 2009; Wang et al., 2015), and the main physical factors—Land Surface Temperature (LST) —controlling the land carbon balance(Anav et al., 2013; Piao et al., 2009). The relationship between LAI, GPP, and environmental factors is complex, as environmental drivers significantly influence LAI dynamics and collectively determine vegetation photosynthetic performance (Zhao et al., 2022; Zhao et al., 2020). Vegetation productivity can be directly estimated using canopy parameters, resulting in a high coupling between GPP and LAI (Walther et al., 2019; Tian et al., 2024).

Both remote sensing and ecosystem models can be effective in investigating large-scale vegetation dynamics (Zhao et al., 2020). Ecosystem models offer an advantage in attributing vegetation growth changes by addressing limitations inherent in traditional statistical models, such as the inability to distinguish correlation from causation and to capture nonlinear relationships (Piao et al., 2006; Zhu et al., 2016). Earth System Models (ESMs), which simulate the physical, chemical, and biological processes and their interactions within the atmosphere, land, ocean, biosphere, and cryosphere (Ziehn et al., 2017; Wu et al., 2019; Bao et al., 2020; Zhang et al., 2020), are critical tools for modeling terrestrial ecosystems (Li et al., 2019a; Lawrence, 2020).

Understanding vegetation responses to past global environmental changes is crucial for improving predictions of future spatiotemporal changes in global vegetation (Zhao et al., 2020). However, numerous studies have evaluated and projected variables simulated by ESMs, revealing a general overestimation trend for model simulations (Anav et al., 2013; Mahowald et al., 2016; Kim et al., 2018; Zeng et al., 2016; Park and Jeong, 2021). Anav et al. (2013) reported that half of the models overestimated LAI trends from 1986 to 2005. Ziehn et al. (2017) found that the ACCESS-ESM1 model overestimated the seasonal mean and peak amplitude of global LAI during historical periods. Zhu et al. (2019) observed that ELMv1-ECA significantly overestimated LAI across most terrestrial surfaces, particularly in the tropics, while underestimating GPP in boreal forest systems and tropical ecosystems. Song et al. (2021) evaluated multiple ESMs and concluded that global LAI overestimation primarily stemmed from overestimated LAI in non-forest areas, with peak LAI in some regions occurring 1–2 months later than observed.

China's carbon sink accounts for approximately 10%-31% of the global terrestrial ecosystem carbon sink (Piao et al., 2022). More accurate simulations of terrestrial ecosystems using CMIP ESM models are highly valuable for both scientific research

and policy development. Currently, CMIP has progressed to its sixth phase (CMIP6). Compared to CMIP5, CMIP6 models feature significantly improved resolution, incorporate more comprehensive biogeochemical processes (Eyring et al., 2016), and consider more plausible future scenarios (O'neill et al., 2016), an advancement that will enhance the potential for simulating terrestrial ecosystems. However, research evaluating the outputs of CMIP6 models remains limited compared to

70   CMIP5 evaluations. In existing evaluation studies, researchers have made significant progress in assessing the performance of ESMs. For example, Zhao et al. (2020) focused on the temporal variability of LAI in CMIP6 models, while Song et al. (2021) evaluated global LAI and tree height performance. However, limitations in spatial resolution and geographical scope have left gaps in the systematic evaluation of ESM simulations for specific regions, such as China. Similarly, Sun et al. (2023) conducted evaluations of LAI, GPP and NPP across Asia but lacked a comprehensive methodology and finer-scale

75   analyses for historical periods. These gaps highlight the need for more detailed and region-specific assessments of CMIP6 ESMs outputs.

Based on the current lack of multi-variable evaluations in ESM assessment studies and the absence of systematic regional evaluations in China, this study utilizes remote sensing data as observational references to evaluate the interannual trends performance of LAI, GPP, NPP, NEP and LST simulated by 12 CMIP6 ESMs (ACCESS-ESM1-5, BCC-CSM2-MR,

80   CanESM5, CESM2-WACCM, EC-Earth3-Veg, INM-CM4-8, INM-CM5-0, IPSL-CM6A-LR, MPI-ESM1-2-HR, MPI-ESM1-2-LR, NorESM2-LM, and NorESM2-MM) in China. A 3×3 sliding window approach is applied spatially to provide a more systematic assessment of trends performance. Finally, remote sensing data of tree cover were used to evaluate the simulated trends of tree distribution patterns.

2 Data and method

85   ### 2.1 The study area

China is situated at the intersection of East and Central Asia, with a geographical range approximately between 18°N and 53°N latitude and 73°E and 135°E longitude. China exhibits diverse topographical features, ranging from expansive plains and hills in the east to high plateaus and mountains in the west, with an overall terrain gradient sloping from west to east. The monsoon region in the east has abundant precipitation and a large temperature difference from north to south, which is

90   suitable for a wide variety of plants to grow, while the arid region in the west has scarce precipitation and sparse vegetation. Against this climatic background, China has developed a diversity of vegetation types ranging from humid broad-leaved evergreen forests to arid desert grasslands, with significant regional differences between seasonal and inter-annual variations in vegetation (Figure 1).
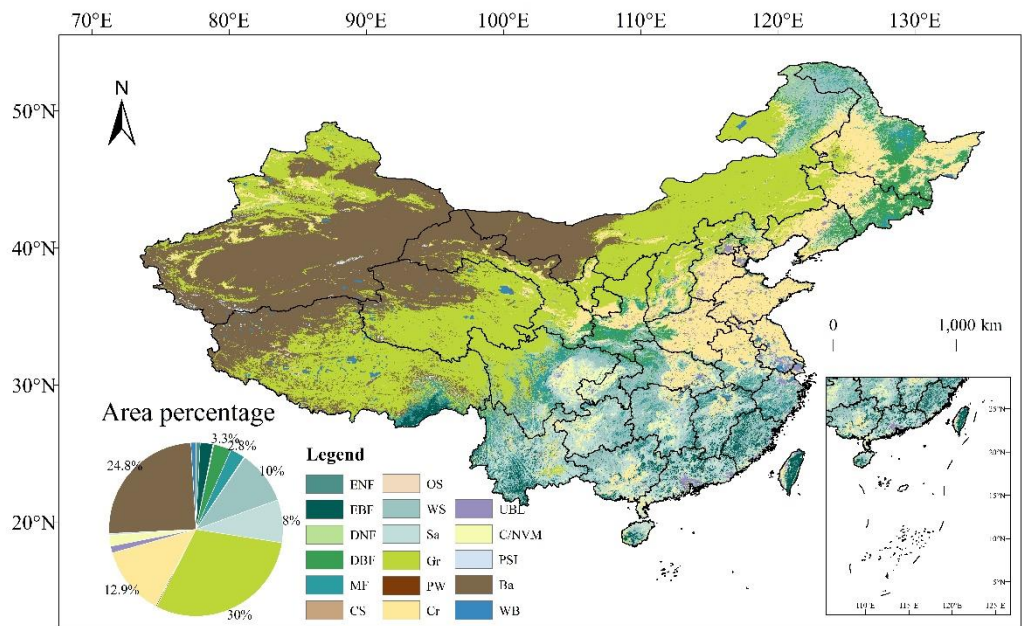
Figure 1: Distribution of vegetation types in China.

2.2 CMIP6 Earth System Models

The Coupled Model Intercomparison Project (CMIP), organized under the auspices of the World Climate Research Programme (WCRP) Working Group on Coupled Modelling (WGCM), which enables the simulation of difficult-to-observe early global coupled climate models by conducting experiments using atmospheric models coupled to the dynamical ocean, a simple land surface, and thermodynamic sea-ice (Meehl et al., 1997). CMIP6 adopts a novel, more coordinated organizational structure (Eyring et al., 2016).

This study utilizes the outputs of LAI and GPP variables from 12 CMIP6 models (ACCESS-ESM1-5, BCC-CSM2-MR, CanESM5, CESM2-WACCM, EC-Earth3-Veg, INM-CM4-8, INM-CM5-0, IPSL-CM6A-LR, MPI-ESM1-2-HR, MPI-ESM1-2-LR, NorESM2-LM, and NorESM2-MM) under five scenarios: historical, SSP1-26, SSP2-45, SSP3-70, and SSP5-85 as the simulated values of the model. To ensure consistency, the outputs from the first realization (r1i1p1) of each model were selected. Monthly data outputs from all models were uniformly resampled to a spatial resolution of $0.5° \times 0.5°$. Additionally, based on whether the models incorporate a dynamic vegetation module, the 12 models were classified into groups with and without dynamic vegetation integration (Table 1).

Table 1: Spatial resolution of selected CMIP6 models and variables information

Model name	Spatial Resolution	LAI	GPP	NPP	NEP	LST	TCF

ACCESS-ESM1-5	1.875°×1.25°	√	√	√	√	√	√
BCC-CSM2-MR	1.125°×1.125°	√	√	√	√	√	×
CanESM5	2.8°×2.8°	√	√	√	√	√	√
CESM2-WACCM	0.9°×1.25°	√	√	√	√	√	√
EC-Earth3-Veg	0.7°×0.7°	√	√	√	√	√	√
INM-CM4-8	2°×1.5°	√	√	√	×	√	√
INM-CM5-0	2°×1.5°	√	√	√	×	√	√
IPSL-CM6A-LR	2.5°×1.27°	√	√	√	√	√	√
MPI-ESM1-2-HR	0.9°×0.9°	√	√	√	×	√	√
MPI-ESM1-2-LR	1.875°×1.875°	√	√	√	√	√	√
NorESM2-LM	2.5°×1.9°	√	√	√	√	√	√
NorESM2-MM	1.25°×0.9°	√	√	√	√	√	√

110   The Multi-Model Ensemble (MME) mean-based approach, an established statistical integration technique, synthesizes outputs from diverse models through weighted averaging(Zeng et al., 2016). This approach combines the true and noise signal, where the inter-model errors is mitigated through model averaging, thereby amplifying underlying true signals.

2.3 Benchmark data

The 2003-2014 LAI data from the reprocessed MODIS product in China (Yuan et al., 2011), were used to validate the LAI
115   outputs of 12 ESMs under the historical scenario. The reprocessed MODIS product refines the original MODIS LAI data (MOD15A2H and MYD15A2H) using modified temporal-spatial filtering (mTSF) technique to fill data gaps and correct low-quality data. Subsequently, the TIMESAT SG filtering technique was applied for post-processing to produce the final product. The reprocessed remote sensing LAI data were then resampled to a 0.5° × 0.5° monthly grid for consistency with the model outputs.

120   The CSIF dataset was used to validate the GPP outputs of the models under the historical scenario. CSIF data are generated by using surface reflectance data from MODIS C6 (MCD43C1) as input (Zhang et al., 2018), trained with daily SIF observations derived from the Orbiting Carbon Observatory-2 using a machine learning algorithm. CSIF effectively captures the seasonal dynamics of satellite-observed SIF and shows a strong correlation with ecosystem GPP, making it a suitable proxy for GPP in vegetation phenology analysis. The CSIF data were aggregated from 4 days to monthly scale data, and the
125   spatial resolution was sampled into a 0.5° × 0.5° grid using mean aggregation.

The NPP dataset was derived from the MOD17A3HGF Version 6 product of MODIS, featuring a spatial resolution of 500 m × 500 m and a temporal resolution of 8 days. Annual NPP values were calculated as the cumulative sum of 8-day Net Photosynthesis (PSN) products for each year. The reprocessed remote sensing NPP data were then resampled to a 0.5° × 0.5° monthly grid for consistency with the model outputs.

130   The NEP dataset was obtained from the reprocessed annual NEP data based on MOD17A3HGF V6 products, provided by
       the Loess Plateau Subcenter of the National Earth System Science Data Center. This dataset retains a spatial resolution of
       500 m × 500 m, the study resampled the remote sensing NEP data to 0.5° × 0.5° monthly grid.

       The LST data were sourced from the Global Daily 0.05° Spatiotemporally Continuous LST dataset hosted by the National
       Tibetan Plateau Data Center. The dataset reconstructs clear-sky LST by applying an Empirical Orthogonal Function (EOF)

135   interpolation method to MODIS Terra/Aqua LST products. It further integrates ERA5-Land climate reanalysis data using a
       Cumulative Distribution Function (CDF) matching method, generating high-quality global spatiotemporally continuous LST
       data under both clear-sky and all-weather conditions. The final product has a spatial resolution of 0.05° × 0.05°, while the
       remote sensing LST dataset were resampled to 0.5° × 0.5° monthly grid in this study.

       Tree cover fraction (TCF) were extracted from the MODIS Terra Vegetation Continuous Fields (VCF) product

140   (https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/science-domain/vegetation-continuous-fields),
       providing global tree cover fraction at a 500 m × 500 m spatial resolution. To consistency with the model output, the remote
       sensing TCF data were resampled to a 0.5° × 0.5° monthly grid.

2.4 Auxiliary data

       The IGBP classification scheme from the MODIS MCD12Q1 product was used to categorize the vegetation in China into

145   four land cover types: forest (Tree Cover), grassland (Grassland), cropland (Crop), and non-vegetated areas (Non-Vegetable).
       Specifically, evergreen coniferous forest, evergreen broadleaf forest, deciduous coniferous forest, deciduous broadleaf forest,
       mixed forest, woody savanna and savanna are combined to be classified as Tree Cover; closed shrubland, open shrubland,
       grassland and permanent wetland are combined to be classified as Grassland; farmland and agricultural land (with natural
       vegetation) are classified as Cropland; and urban and built-up land, permanent snow and ice, unutilized land, and water are

150   combined to be classified as Non-Vegetable areas. The specific classification details are summarized in Table 2.

Table 2: Summary of the classification of vegetation and land use types in China.

Name	Vegetation or land use type	Classification type	area proportion (%)
ENF	Evergreen Needleleaf Forests	Tree Cover	0.77
EBF	Evergreen Broadleaf Forests	Tree Cover	2.37
DNF	Deciduous Needleleaf Forests	Tree Cover	0.15
DBF	Deciduous Broadleaf Forests	Tree Cover	3.29
MF	Mixed Forests	Tree Cover	2.75
CS	Closed Shrublands	Grassland	0.04
OS	Open Shrublands	Grassland	0.08
WS	Woody Savannas	Tree Cover	10.01
Sa	Savannas	Tree Cover	7.96

Gr	Grasslands	Grassland	30.04
PW	Permanent Wetlands	Grassland	0.21
Cr	Croplands	Crop	12.92
UBL	Urban and Built-up Lands	Non Vegetable	1.25
C/NVM	Cropland/Natural Vegetation Mosaics	Crop	2.02
PSI	Permanent Snow and Ice	Non Vegetable	0.33
Ba	Barren	Non Vegetable	24.82
WB	Water Bodies	Non Vegetable	0.99

2.5 Analysis

The Theil-Sen Median method was employed to estimate the interannual trends of observed and simulated LAI, GPP, NPP, NEP and LST from 2003 to 2014, while the Mann-Kendall (MK) test was applied to assess the statistical significance of these trends. Additionally, the performance of each model in simulating the interannual LAI, GPP, NPP, NEP and LST time series was quantified using the standard deviation (SD) of interannual time series, correlation coefficient (r), and root mean square error (RMSE).

Observed LAI and CSIF trends were used to validate the spatial trends of simulated LAI and GPP, providing a comprehensive spatial assessment of model performance. At the pixel scale, a 3×3 sliding window approach was applied to extract paired pixel value sequences from model simulations and satellite observations. These paired sequences were statistically compared using a paired t-test (also known as a dependent or correlated t-test). This method compares the means and standard deviations of two related groups to determine whether significant differences exist between them. The hypotheses for the paired t-test are formulated as follows: H_0: $m=0$ (null hypothesis, where m is the mean of differences between paired samples); H_1: $m \neq 0$ (alternative hypothesis).

$$t = \frac{m}{S/\sqrt{n}} \tag{1}$$

$$S = \sqrt{\frac{\sum_{i=1}^{n}(d_i - m)^2}{n-1}} \tag{2}$$

$$m = \frac{\sum_{i=1}^{n} d_i}{n} \tag{3}$$

Here, $i=1,2,\dots,n$. Where m represents the mean of paired sample differences, S denotes the standard deviation of these differences, and n is the number of paired samples. The critical t-value is determined by degrees of freedom (df) and significance level (α). If the calculated t-value is less than the critical value at α-level (P), this indicates $P>\alpha$, leading to

acceptance of H_0 and rejection of H_1 at the α significance level. Conversely, if the calculated t-value exceeds the critical value, H_0 is rejected in favor of H_1.

To enable a comparison between the trends of CSIF and GPP, it is necessary to normalize the annual-scale CSIF data and the GPP outputs from the ESMs. Normalization ensures that both datasets are brought to a comparable scale, facilitating a more accurate and meaningful trend analysis.

$$X = \frac{X_i - mean}{std} \tag{4}$$

Where X is the annual normalized CSIF or GPP data, X_i is the annual unnormalized data, *mean* is the average of the multi-year annual data, and *std* is the standard deviation of the multi-year annual data.

3 Results

3.1 Trend of observed variables in China

From the spatial trends of MODIS LAI across China from 2003 to 2014 (Fig. 2a), 82.60% and 17.40% of the study area exhibited increasing and decreasing LAI trends, respectively. Among these, areas with statistically significant increasing and decreasing trends accounted for 44.10% and 7.24%, respectively. Spatially, declining LAI trends were primarily concentrated in grassland and cultivated areas of northwestern Xinjiang, central Inner Mongolia grasslands, and parts of the Tibetan Plateau. In contrast, significant increases in LAI were predominantly observed in southern forested regions, grassland areas of the Loess Plateau, and forest-cultivated zones of northeastern China.
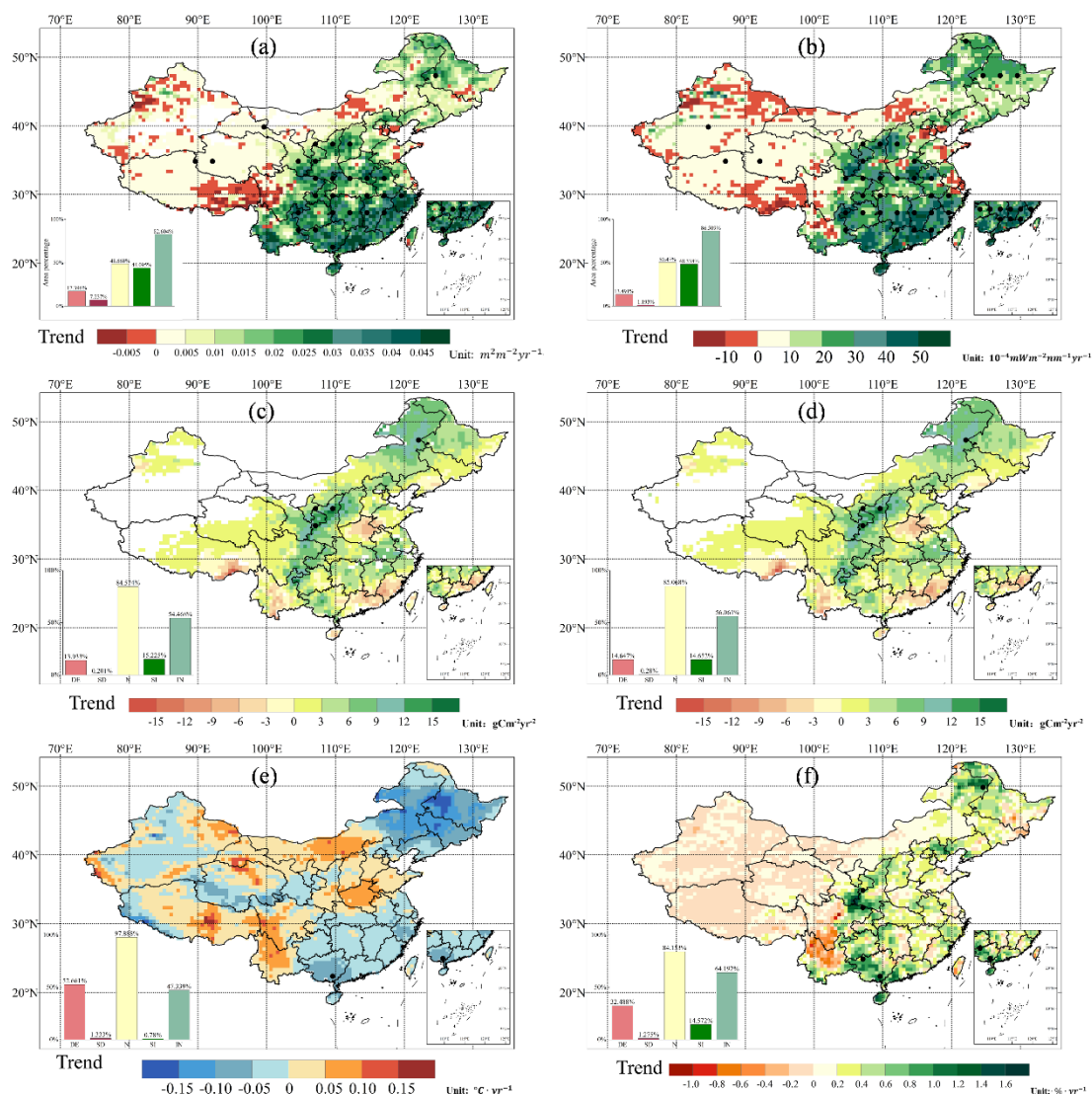
Figure 2: The observed annual (a) MODIS LAI, (b) CSIF, (c) MODIS NPP, (d) MODIS NEP, (e) MODIS LST and (f) MODIS TCF trend during 2003~2014 in China. Pink color represents the percentage of area of decreasing regions (DE), red color
190 **represents the percentage of area of significantly decreasing regions (SD), yellow color represents the percentage of area of regions with no significant change (N), green color represents the percentage of area of significantly increasing regions (SI), and light green color represents the percentage of area of increasing regions (IN). The dot indicated the significant trend (p<0.05).**

Figure 2b illustrates the spatial trends of CSIF in China during 2003–2014. Overall, 86.50% and 13.49% of the region displayed increasing and decreasing CSIF trends, respectively, with significant increases and decreases covering 48.34% and
195 1.19% of the area. The trends in CSIF during 2003–2014 exhibited similar patterns to those of LAI, particularly in southern forested regions, northeastern China, and the Loess Plateau, where both variables showed significant increases. However, CSIF exhibited more significant rising trends in the North China Plain compared to LAI. Conversely, while declining trends

in CSIF were observed in parts of Xinjiang, Inner Mongolia, and the Tibetan Plateau—consistent with LAI patterns—CSIF reductions were more prominent in central Yunnan, the Yangtze River Delta, and the Beijing-Tianjin-Hebei region.

200   The spatial trends of MODIS NPP in China during 2003–2014 reveal that most regions exhibited no statistically significant changes (Fig. 2c), with 84.57% of the area showing non-significant variability. Areas with increasing and decreasing trends accounted for 54.47% and 13.93% of the total area, respectively, while statistically significant increases and decreases covered 15.23% and 0.20% of the region. Spatially, reductions in remote sensing-derived NPP were primarily distributed in the Central Plains region, Guangdong and Fujian provinces, central Yunnan, and southern Tibet. In contrast, significant

205   increases in NPP were notably observed in northern Shaanxi and Ningxia regions, as well as parts of northeastern China.

The spatial trends of MODIS NEP in China during 2003–2014 (Fig. 2d) closely resembled the spatial patterns of MODIS NPP. Across China, 56.06% and 14.65% of the area exhibited increasing and decreasing NEP trends, respectively. Among these, 85.07% of the region showed no statistically significant changes, while significant increases and decreases accounted for 14.65% and 0.28% of the total area. Similar to the spatial distribution of remote sensing NPP trends, significant NEP

210   increases were observed in northern Shaanxi and Ningxia regions, as well as parts of northeastern China. Conversely, declining trends in NEP were concentrated in the North China Plain, Guangdong and Fujian provinces, central Yunnan, and southern Tibet.

Figure 2e illustrates the spatial trends of MODIS LST across China during 2003–2019. Overall, 47.34% and 52.66% of the study area exhibited increasing and decreasing LST trends, respectively. However, only 0.78% and 1.33% of the area

215   showed significant increases and decreases, while 97.89% of the region displayed no statistically significant changes. Spatially, rising LST trends were concentrated in central-western Inner Mongolia, eastern Xinjiang, the Gansu-Qinghai region, the Sichuan-Tibet Plateau, Yunnan Province, and the North China Plain. Conversely, declining LST trends were primarily observed in northeastern China and southern China.

To better investigate the promoting effects of human activities, such as afforestation and the Grain-for-Green Program, on

220   vegetation growth, we analysed the spatial trends of MODIS TCF changes in China during 2003–2014 (Fig. 2f). The results indicate that 64.19% and 32.49% of the study area exhibited increasing and decreasing TCF trends, respectively. Among these, statistically significant increases and decreases accounted for 14.57% and 1.28% of the total area, while 84.15% showed no significant trend. Most tree-abundant regions in eastern and central China displayed TCF gains. In contrast, declining TCF trends were observed in western China, where natural conditions are unsuitable for tree growth and land cover is predominantly grassland or non-vegetated areas. Significant TCF increases were concentrated in the Shaanxi-

225   Gansu-Ningxia region, the Greater Khingan Mountains region in northeastern China, and Guizhou, Guangdong, and Guangxi provinces.

3.2 Evaluating the overall trends of CMIP6 variables

As shown in the bar chart comparing observed LAI trends with MME simulated LAI (Fig. 3a), the MODIS LAI exhibited a

230   trend of 0.11 m²·m⁻²·yr⁻¹, which falls within the MME uncertainty range (0.0060 ± 0.0058 m²·m⁻²·yr⁻¹). This indicates that,

overall, the models are capable of capturing the observed LAI trend. However, it is worth noting that the trend values simulated by individual models vary significantly, ranging from -0.0021 yr⁻¹ (INM-CM5-0) to 0.019 yr⁻¹ (NorESM2-MM), contributing to substantial uncertainties in trends of simulated LAI.

A similar analysis was conducted for the normalized interannual trends of observed CSIF and simulated GPP in China

235 during 2003–2014 (Fig. 3c). The observed CSIF exhibited a significant trend of 0.13 yr⁻¹, which substantially exceeded all model simulations, including the MME. Simulated GPP trends ranged from -0.020 yr⁻¹ (INM-CM5-0) to 0.10 yr⁻¹ (MME). These results demonstrate a systematic underestimation of GPP trends in China by DGVMs.
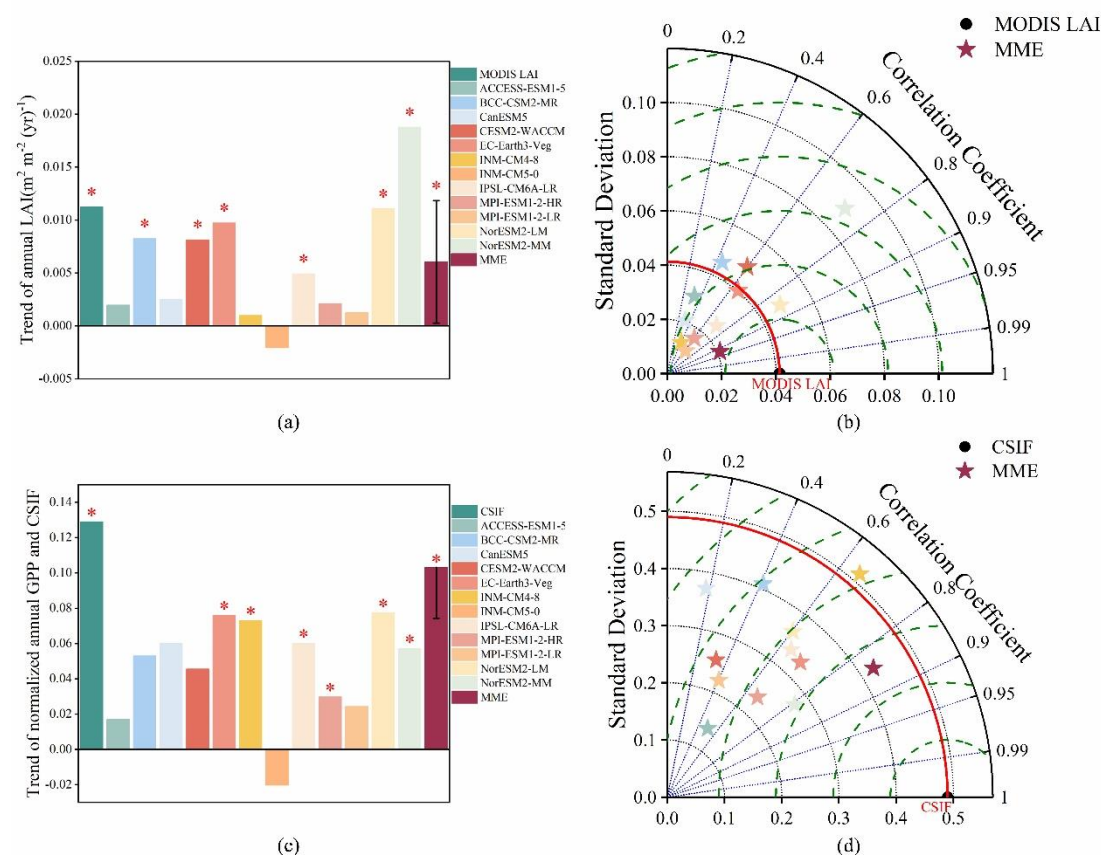
Figure 3: Overall annual average (a) LAI and (c) GPP trends in China during 2003-2014. The asterisk (*) indicated the significant
240 **trend (p<0.05). The Taylor diagrams compare the remotely-sensed and model-simulated global annual mean (b) LAI and (d) GPP for the historical period (2003-2014). The standard deviation shows the interannual variability of the observed and modeled LAI. The dashed green lines show centered root mean square difference (RMSD) between model simulations and satellite observations.**

The Taylor diagram (Fig. 3b, 3d) is employed to evaluate and compare the interannual variations of LAI and GPP among individual models, the MME, and observations. According to the reference observational data, the standard deviation (SD) of
245 the interannual time series for LAI and CSIF over the Chinese region during 2003–2014 is 0.041 and 0.49, respectively. The SD of ESM-simulated LAI ranges from 0.011 (MPI-ESM1-2-LR) to 0.089 (NorESM2-MM), while that of GPP ranges from

0.139 (ACCESS-ESM1-5) to 0.515 (INM-CM4-8). The MME-simulated SD for LAI and GPP are 0.021 and 0.425, respectively. The root mean square deviations (RMSDs) of the models for LAI range from 0.024 (NorESM2-LM) to 0.063 (NorESM2-MM), and for GPP, from 0.334 (EC-Earth3-Veg) to 0.697 (INM-CM5-0). The MME-simulated RMSD for LAI and GPP are 0.022 and 0.249, respectively. Generally, the MME-simulated SD and RMSD for both LAI and GPP are smaller than those of most individual models. The correlation coefficients between the ESM-simulated interannual LAI time series and the observed LAI range from -0.409 (INM-CM5-0) to 0.855 (NorESM2-LM). The correlation coefficient for the MME-simulated LAI is 0.922, exceeding that of any individual model. Similarly, the correlation coefficients between the ESM-simulated GPP and observed CSIF time series range from -0.207 (INM-CM5-0) to 0.809 (NorESM2-MM), with the MME-simulated GPP showing a correlation coefficient of 0.847, which is higher than that of all individual models.

Figure 4a illustrates the observed NPP exhibited a trend of 2.70 g C·m⁻²·yr⁻¹, which lies within the uncertainty range of the MME trend (1.19 ± 1.53 g C·m⁻²·yr⁻¹), indicating that models broadly captured the direction of NPP changes (Fig. 6a). However, three models—ACCESS-ESM1-5 (-0.49 g C·m⁻²·yr⁻¹), INM-CM5-0 (-1.42 g C·m⁻²·yr⁻¹), and MPI-ESM1-2-LR (-0.27 g C·m⁻²·yr⁻¹)—simulated declining NPP trends. In contrast, MODIS NPP showed a statistically significant increasing trend during 2003–2014, whereas only INM-CM4-8, NorESM2-LM, and the MME produced significant positive trends among the models.
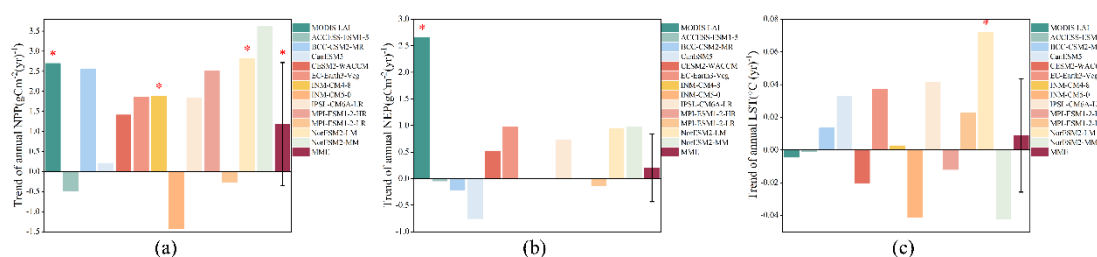

Figure 4: Overall annual average (a) NPP, (b) NEP and (c) LST trends in China during 2003~2014. The asterisk (*) indicated the significant trend (p<0.05).

The observed NEP exhibited a statistically significant increasing trend (2.66 g C·m⁻²·yr⁻¹), whereas simulated NEP trends were systematically lower than observations (Fig. 4b). The MME simulated NEP trend of 0.20 ± 0.64 g C·m⁻²·yr⁻¹—substantially below the observed value—with exceptionally large model simulations uncertainties. Specifically, ACCESS-ESM1-5 (-0.05 g C·m⁻²·yr⁻¹), BCC-CSM2-MR (-0.21 g C·m⁻²·yr⁻¹), CanESM5 (-0.76 g C·m⁻²·yr⁻¹), and MPI-ESM2-LR (-0.14 g C·m⁻²·yr⁻¹) simulated declining NEP trends, highlighting the models' underestimation of carbon sink intensification.

The observed LST exhibited declining trend of -0.0045 °C·m⁻²·yr⁻¹ in China during 2003–2014 (Fig. 4c). which lies within the uncertainty range of the MME simulated a positive LST trend (0.0088 ± 0.035 °C·m⁻²·yr⁻¹). Notably, only five models—ACCESS-ESM1-5, CESM2-WACCM, INM-CM5-0, MPI-ESM1-2-HR, and NorESM2-MM—captured negative LST trends, whereas all others simulated positive LST trend. The NorESM2-LM model produced a statistically significant positive LST trend.

275 **3.3 Assessment of spatial patterns and trends in CMIP6 variables**

From a spatial perspective, a comparison between simulated and observed LAI trends reveals that most models exhibit significant overestimation and underestimation in China (Fig. 5). The MME overestimated, underestimated, and simulated well with observed LAI trends in 21.71%, 46.29%, and 32.00% in China, respectively. Specifically, the MME predominantly underestimated LAI trends in regions with statistically significant observed increases—particularly southern China, the

280 Loess Plateau, and northeastern China—while overestimating trends in areas with declining or non-significant LAI changes (e.g., northwestern Xinjiang, the Tibetan Plateau, and parts of the North China Plain). Among individual models, EC-Earth3-Veg exhibited relatively better performance, but even this model aligns with 40.36% of its simulated LAI trends matching observations (Fig. 5f). However, most models showed dominant underestimation patterns: CanESM5 (62.97%), INM-CM4-8 (60.09%), and INM-CM5-0 (63.50%) had the largest proportions of underestimated LAI trends.
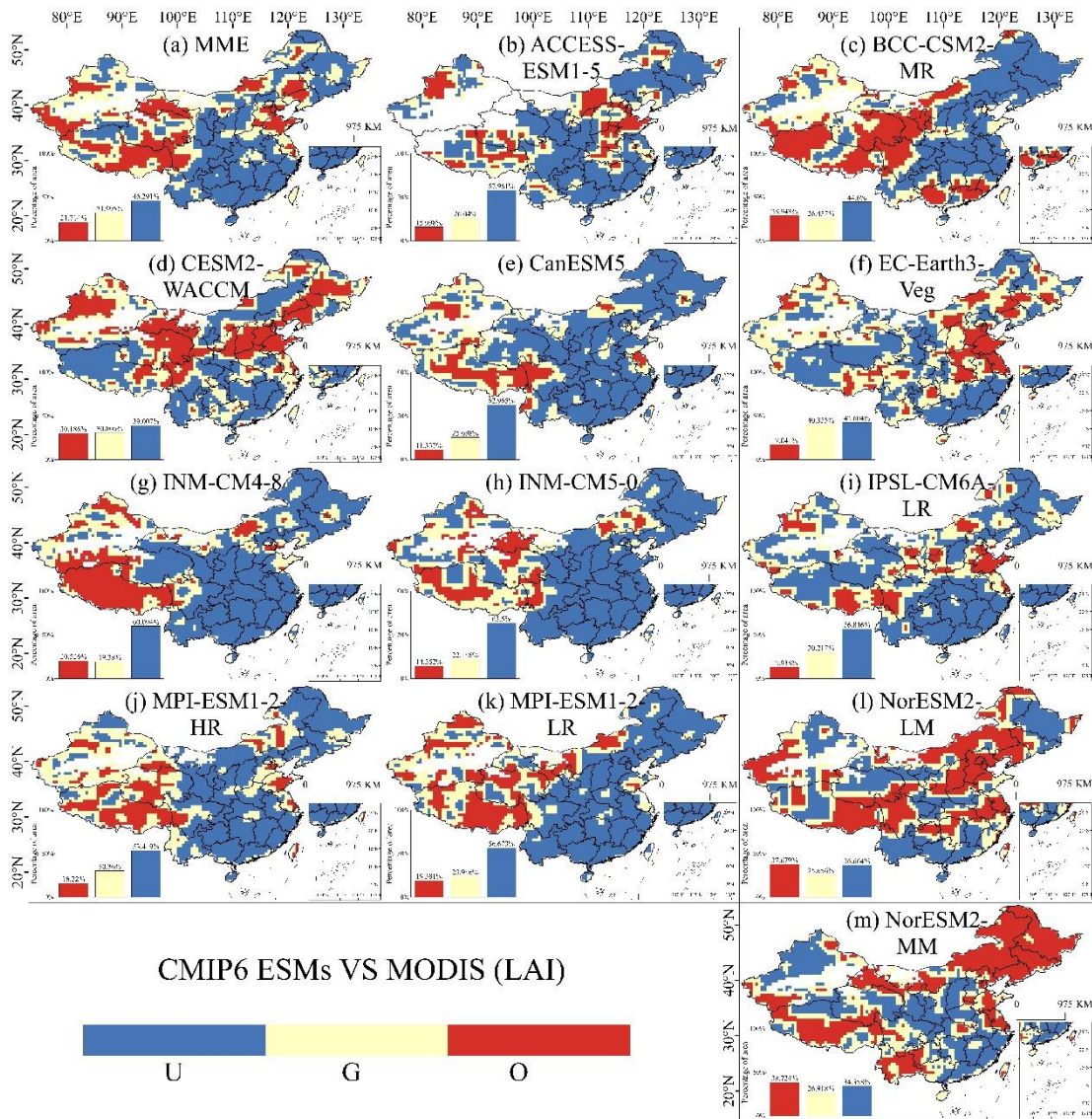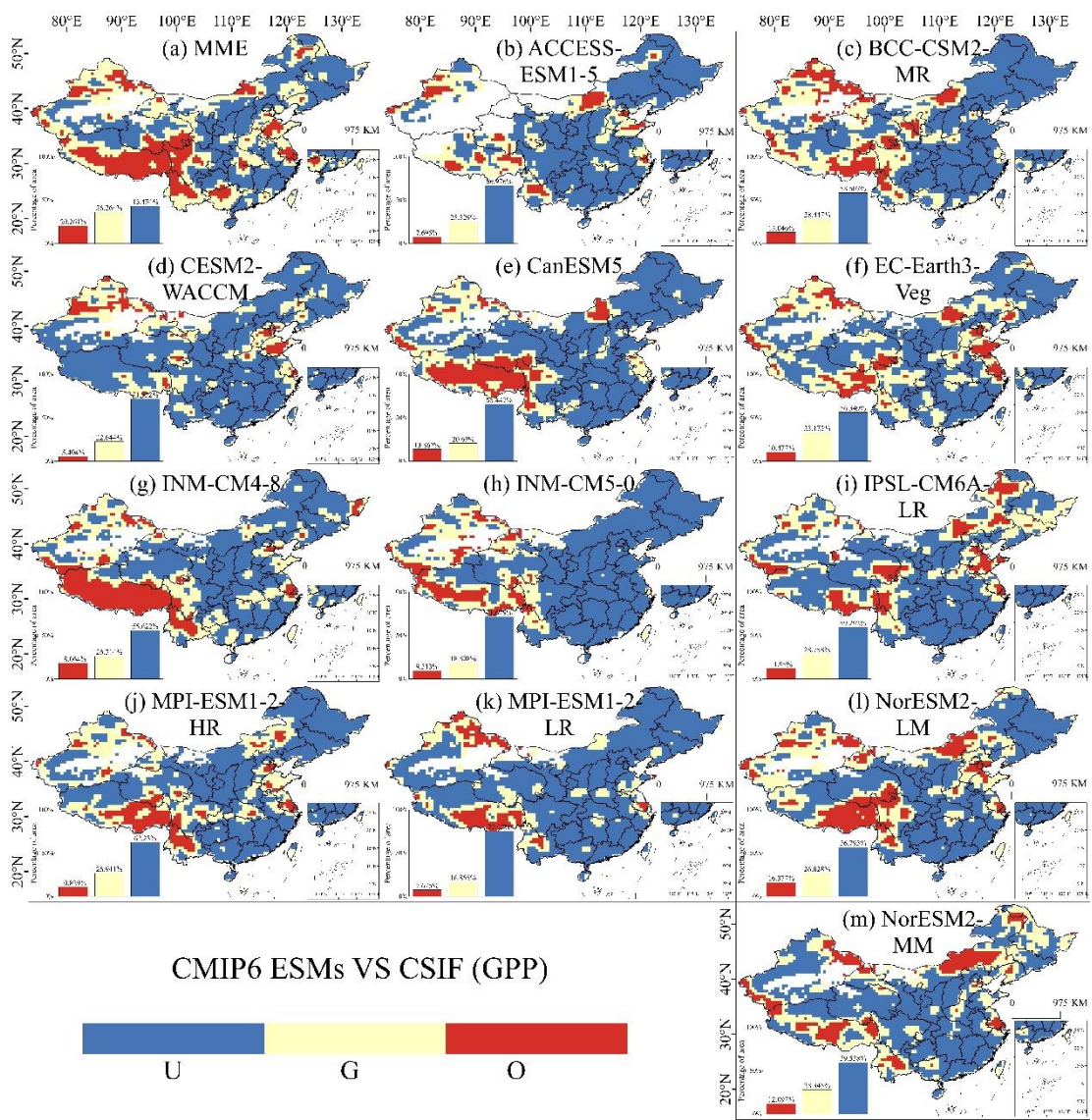
Figure 5: Evaluation of the LAI trend performance of the CMIP6 ESMs from 2003 to 2014 in China. A paired t-test with a sample size of 9 was conducted using a 3 × 3 sliding window to determine whether the model simulated a trend that was not significantly different from the observed data (Good), significantly smaller than the observed data (Underestimate), or significantly larger than the observed data (Overestimate). U, G, and O represent 'Underestimate ', 'Good', and 'Overestimate', respectively.

The spatial distribution of normalized GPP trend discrepancies between simulated and observed CSIF exhibited similar patterns to that of the model LAI performance (Fig. 6). For the MME, overestimated, underestimated, and simulated well GPP trends accounted for 20.26%, 43.47%, and 36.26% in China, respectively (Fig. 6a). Regions with overestimated GPP trends were predominantly distributed in Yunnan Province, the Sichuan-Tibet Plateau, and northwestern border areas. Among individual models, INM-CM4-8 exhibited the largest proportion of overestimated GPP trends (18.06%), alongside

14

295 55.62% underestimated areas (Fig. 6g). In contrast, MPI-ESM1-2-LR showed extreme underestimation biases, with 75.47%

of its simulated GPP trends being lower than observations—the highest underestimation proportion among all models—

while only 7.68% of areas were overestimated (Fig. 6k).

Figure 6: Evaluation of the normalized GPP trend performance of the CMIP6 ESMs from 2003 to 2014 in China. A paired t-test
300 **with a sample size of 9 was conducted using a 3 × 3 sliding window to determine whether the model simulated a trend that was not**
significantly different from the observed data (Good), significantly smaller than the observed data (Underestimate), or
significantly larger than the observed data (Overestimate). U, G, and O represent 'Underestimate ', 'Good', and 'Overestimate',
respectively.

The spatial patterns of discrepancies between simulated and MODIS NPP trends during 2003–2014 are shown in Fig. 7.

305 Simulated NPP trends were generally overestimated in southwestern and southern China but underestimated across most

15

other regions (Fig. 7). For the MME, overestimated, underestimated, and simulated well NPP trends occupied 28.62%, 49.81%, and 21.57% of the study area, respectively. Overestimated regions were predominantly concentrated in the Central Plains, Sichuan-Tibet Plateau, and Yunnan Province (Fig. 7a). Notably, the NorESM2-LM model showed the highest proportion of overestimated NPP trends (47.46%) among all models, with underestimated and simulated well areas covering 33.84% and 18.70%, respectively. Its overestimations spanned southwestern China, southern China, the North China Plain, the middle Yangtze River Basin, and parts of northeastern China (Fig. 7l). Two other models—MPI-ESM1-2-HR and NorESM2-MM—also exhibited substantial overestimation (42.88% and 46.07%, respectively), while their underestimation proportions reached 29.86% and 28.30%. In contrast, CanESM5 and INM-CM5-0 displayed most underestimation biases, with 62.27% and 69.85% of their simulated NPP trends falling below observations.
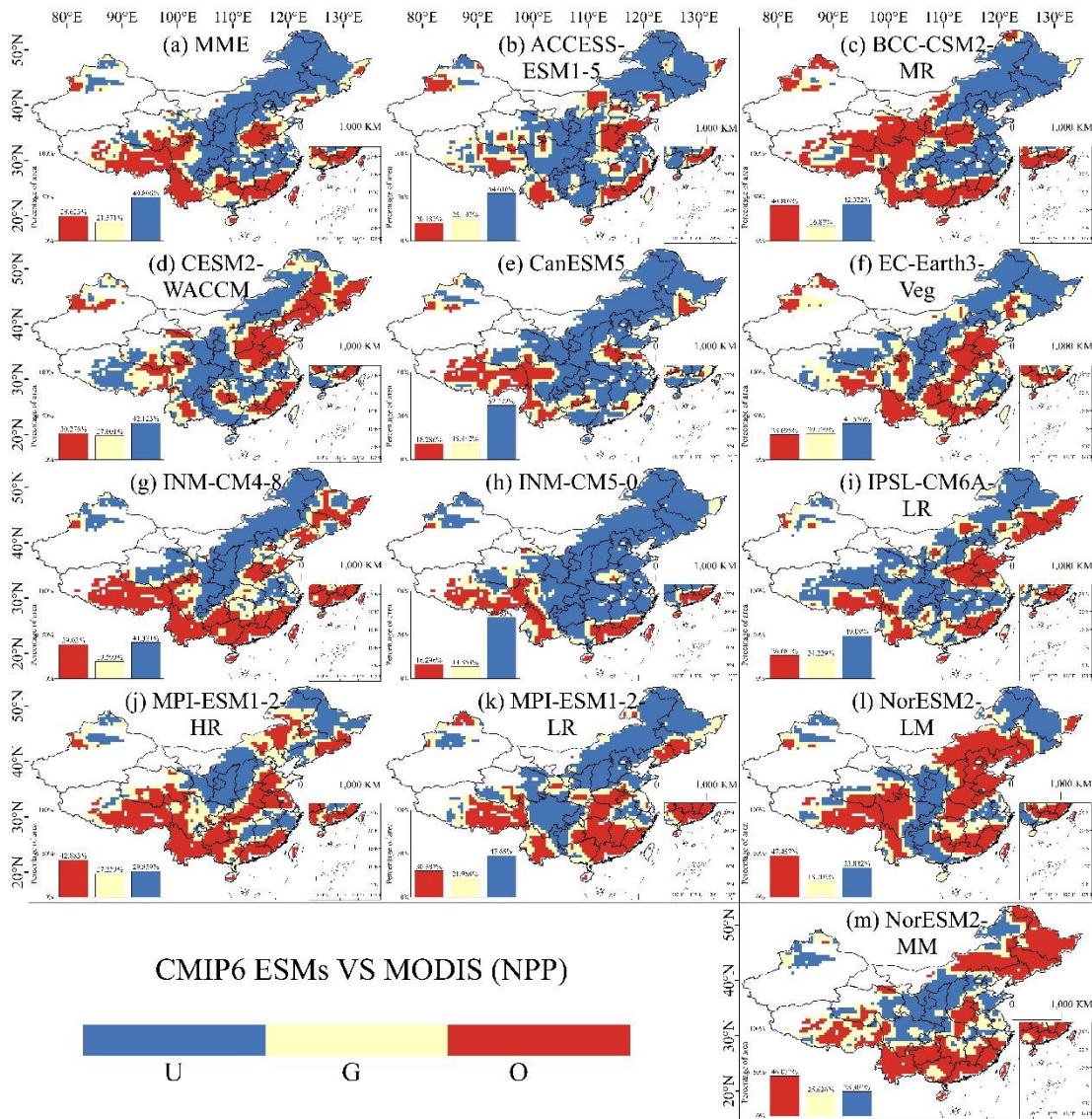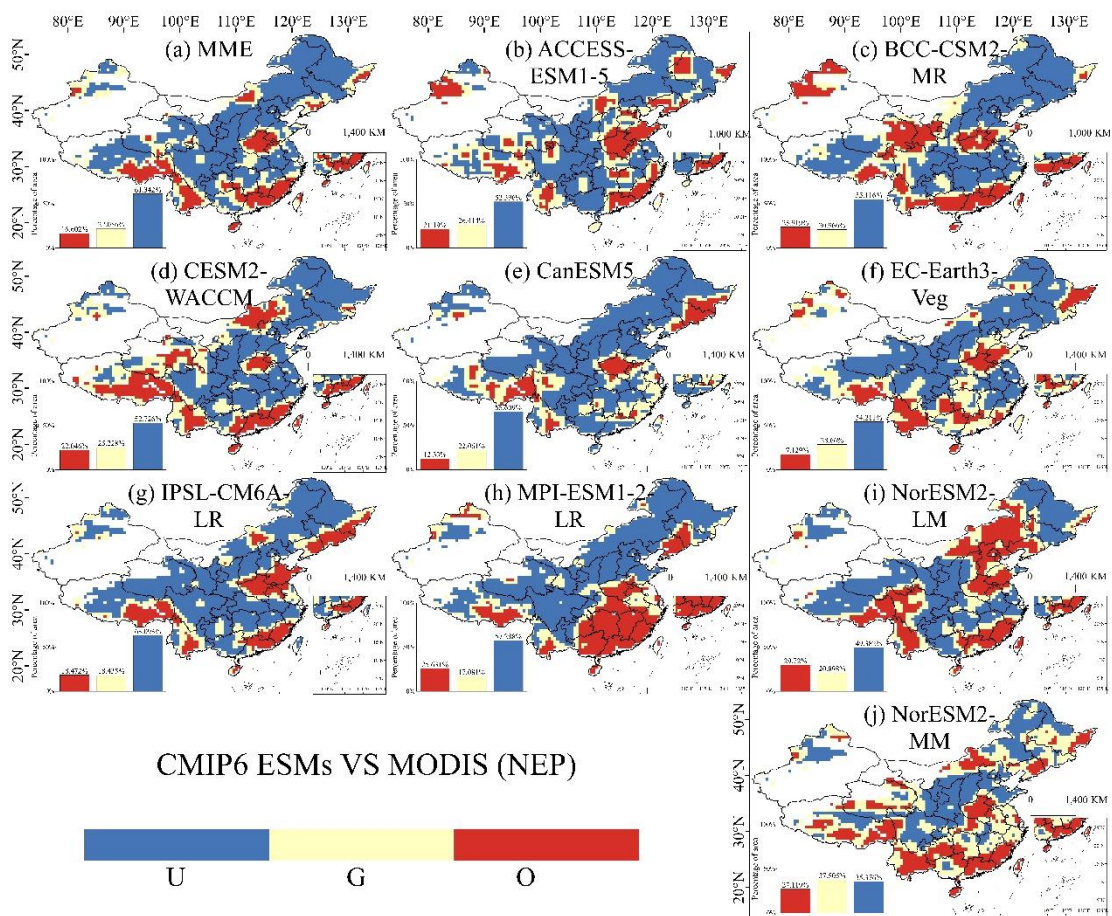
Figure 7: Evaluation of the NPP trend performance of the CMIP6 ESMs from 2003 to 2014 in China. A paired t-test with a sample size of 9 was conducted using a 3 × 3 sliding window to determine whether the model simulated a trend that was not significantly different from the observed data (Good), significantly smaller than the observed data (Underestimate), or significantly larger than the observed data (Overestimate). U, G, and O represent 'Underestimate ', 'Good', and 'Overestimate', respectively.

320   Figure 8 displays spatial discrepancies between simulated and MODIS NEP trends across China during 2003–2014. The MME underestimated, overestimated, and simulated well NEP trends in 61.34%, 16.60%, and 22.06% in China, respectively (Fig. 8a), with spatial patterns similar to that of the simulated NPP trend spatial performance, indicating systemic uncertainties in coupled carbon flux simulations. Among individual models, NorESM2-LM exhibited the highest proportion of overestimated NEP trends (29.72%), alongside 49.38% underestimation and 20.90% simulated well (Fig. 8i), while

17

NorESM2-MM showed the lowest underestimation rate (35.38%) with 27.12% overestimation and 37.51% simulated well (Fig. 8j). In contrast, all other models underestimated NEP trends in over 50% of the region, with CanESM5 displaying the most severe underestimation bias (65.61% underestimated, 12.33% overestimated) (Fig. 8e).
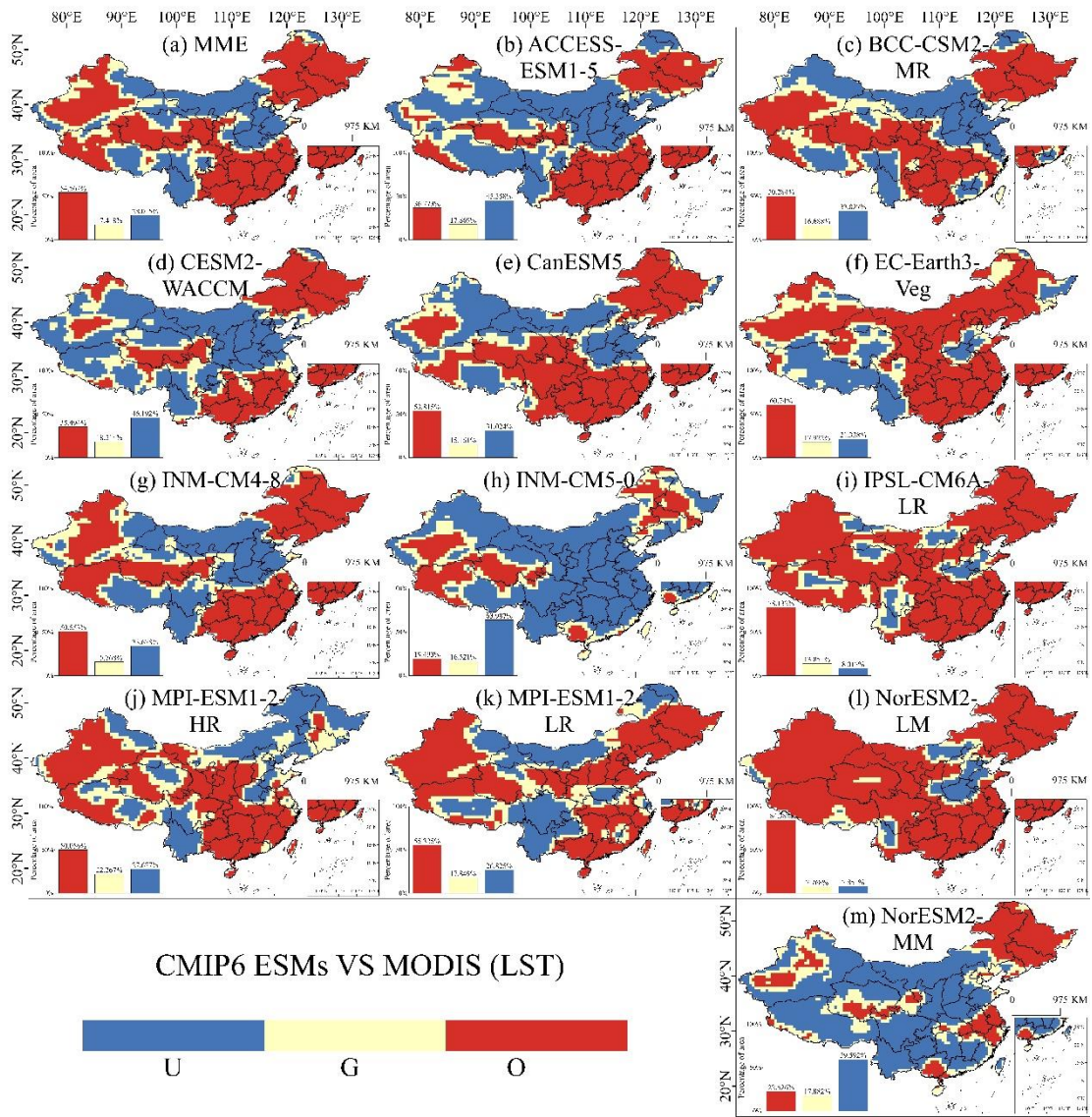

Figure 8: Evaluation of the NEP trend performance of the CMIP6 ESMs from 2003 to 2014 in China. A paired t-test with a sample size of 9 was conducted using a 3 × 3 sliding window to determine whether the model simulated a trend that was not significantly different from the observed data (Good), significantly smaller than the observed data (Underestimate), or significantly larger than the observed data (Overestimate). U, G, and O represent 'Underestimate ', 'Good', and 'Overestimate', respectively.

Figure 9 illustrates spatial differences between simulated and observed LST trends in China during 2003–2014. Most models overestimated LST trends, with the MME overestimating, underestimating, and simulating well observations in 54.57%, 28.02%, and 17.42% of the study area, respectively (Fig. 9a). Spatially, the MME overestimated LST trends in southern China, northeastern China, Xinjiang, the Tibetan Plateau, and the Shaanxi-Gansu region compared to observations. While some models proportion of underestimated LST trend—including ACCESS-ESM1-5 (45.36%), CESM2-WACCM (46.19%), INM-CM5-0 (63.99%), and NorESM2-MM (59.59%)—showed substantial underestimation biases, the majority of models still exhibited larger overestimated areas. Notably, NorESM2-LM displayed the most extreme overestimation, with 84.38%

340 of its simulated LST trends exceeding observations, except in localized underestimations over the North China Plain, central Inner Mongolia, and parts of Yunnan (Fig. 9l).

Figure 9: Evaluation of the LST trend performance of the CMIP6 ESMs from 2003 to 2014 in China. A paired t-test with a sample size of 9 was conducted using a 3 × 3 sliding window to determine whether the model simulated a trend that was not significantly different from the observed data (Good), significantly smaller than the observed data (Underestimate), or significantly larger than the observed data (Overestimate). U, G, and O represent 'Underestimate ', 'Good', and 'Overestimate', respectively.

19

4 Discussion

4.1 ESM captures CMIP6 variables trends in China

Previous studies have found that, whether for CMIP5 or CMIP6, models have not effectively captured the long-term trends
of vegetation (Anav et al., 2013; Song et al., 2021), Overestimations of LAI, GPP, and NPP are more common in the mid-latitude regions of Asia (Sun et al., 2023). This study also reveals that the models have misrepresented the trends of LAI and
GPP in the China region, with significant underestimation of LAI and GPP growth in the forested areas of central and
southern China, and overestimation of LAI and GPP trends in grassland areas of eastern Inner Mongolia and the Tibetan
Plateau, where the growth or decline of LAI and GPP is less significant. Ecosystem models suggest that the global greening
of vegetation is mainly driven by the increase in atmospheric CO_2 concentrations, while climate change and land-use
changes are strong regional drivers of vegetation dynamics (Chen et al., 2019; Zhu et al., 2016). The misrepresentation of
LAI and GPP trends in the models can be attributed to unreasonable parameter settings and missing mechanisms (Song et al.,
2021). The overestimation of the growing season length (GSL) (Verger et al., 2016) in the models is a key factor
contributing to the widespread overestimation of LAI in northern and temperate regions. Additionally, in the models, the
reduction in water limitation mechanisms results in the overestimation of LAI in non-forest areas (Song et al., 2021). The
overestimation of maximum carboxylation rates in the models leads to excessive biomass allocation to leaves, further
overestimating LAI in some regions, compounded by missing mechanisms (Li et al., 2019a). The lack of adequate
mechanisms for soil moisture and nutrient limitations in the models results in an overestimation of the positive drivers for
increased LAI caused by elevated atmospheric CO_2 concentrations, temperature rise, and woody plant invasions (Huang et
al., 2017; Reich et al., 2014; Wang et al., 2014).

Despite substantial progress in understanding vegetation autotrophic respiration through field experiments and observational
studies, significant challenges remain in simulating respiratory processes and their responses to environmental changes due
to inherent physiological complexities (Smith and Dukes, 2013; Yuan et al., 2024). Empirical studies suggest that the ratio of
NPP to GPP often converges toward a constant value across regions, vegetation adapts photorespiration to changes in
photosynthesis through self-regulation (Gifford, 1995; Dewar et al., 1998; Collalti and Prentice, 2019). However, respiratory
processes are inherently harder to quantify than photosynthesis, leading models to rely heavily on empirical functions for
simulating NPP and NEP. Such parameterizations often fail to capture real-world vegetation respiratory acclimation. Spatial
discrepancies in ESM simulated NPP trends (Fig. 7) mirror those of GPP trends (Fig. 6), with systematic underestimations in
northeastern forested areas, the Loess Plateau, Inner Mongolian grasslands, and the middle-lower Yangtze River Basin,
alongside overestimations over Tibetan Plateau grasslands. These biases indicate that ESMs primarily misrepresent trends in
photosynthetic capacity, which leads to misestimation in simulated NPP. Satellite observations reveal pronounced NPP
declines in Guangdong and Fujian provinces during 2003–2014 (Fig. 2c), yet models paradoxically overestimate NPP trends
in these regions. Critically, the NPP and GPP ratio is not static but varies across ecosystems and responds to environmental
drivers (Zhang et al., 2009; Amthor, 2000; Delucia et al., 2007). For instance, temperature, precipitation, and rising CO_2

20

380   levels fertilization can enhance plant carbon-use efficiency by reducing photorespiration, thereby increasing NPP (Slot and Kitajima, 2015; Schuur et al., 2001; Aspinwall et al., 2017), leads to changes of the ratio of NPP and GPP (Zhang et al., 2009). Schuur et al. (2001) demonstrated that excessive rainfall in warm-humid regions (e.g., southern China) may leach nutrients, reduce light availability, and impair soil oxygen diffusion—all suppressing NPP. Current ESMs inadequately represent these nuanced respiratory responses to environmental, particularly in regions like Guangdong and Fujian.

385   ESMs face critical limitations in representing soil carbon-atmosphere interactions and respiratory mechanisms, largely due to insufficient understanding of large-scale vegetation autoregulatory respiration and soil heterotrophic respiration processes (Zhao et al., 2019; Tang et al., 2020). These models often lack mechanistic descriptions of soil respiration (Tang et al., 2020) and struggle to simulate soil carbon-climate feedbacks (Lehmann and Kleber, 2015). The spatial discrepancies between simulated and observed NEP trends in China (Fig. 8) closely resemble those of NPP trends (Fig. 7), as NEP—calculated as

390   the difference between NPP and soil heterotrophic respiration—inherits biases from both components. Models fail to capture regional heterogeneity in soil respiration, and misestimation in simulated NPP (e.g., underestimation in afforestation zones or overestimation in grasslands) propagate directly into NEP biases. For instance, systematic underestimation of NPP in northeastern forests and the Loess Plateau (Fig. 8), combined with oversimplified soil respiration parameterizations, leads to misestimation in simulated NEP. This highlights a fundamental issue: current models cannot disentangle uncertainties in

395   photosynthesis and respiration, nor represent their spatially divergent responses to environmental drivers. These limitations underscore the urgent need for mechanistic advancements in simulating soil-vegetation-atmosphere continuum dynamics.

In southern China and northeastern forested regions, simulated LST trends were significantly higher than observed values (Fig. 9), coinciding with systematic underestimation of LAI trends in these areas (Fig. 5). This contradiction highlights a critical model bias: increased LAI in forested ecosystems reduces LST through enhanced shortwave radiation interception,

400   modified albedo, and altered aerodynamic and surface resistances (Li et al., 2015; Li et al., 2019b). Vegetation growth typically cools surface temperatures by improving canopy shading and evapotranspiration efficiency—mechanisms poorly represented in current parameterizations. The models' failure to capture LAI increases thus leads to simulated LST warming trends, underscoring the necessity for improved representation of biophysical feedbacks in land surface.

4.2 Potential anthropogenic impact for ESM inaccuracy

405   A series of ecological engineering initiatives, such as afforestation, reforestation, and grassland restoration in southern China and the Loess Plateau, have greatly promoted vegetation growth. However, the models may fail to capture the effects of these human activities. ESMs likely fail to adequately represent these anthropogenic impacts.

Figures 10 and Figures 11 compare observed and simulated TCF trends at both overall and spatial scales during 2003–2014. The observed TCF exhibited a significant increasing trend of 0.14% yr^{-1} (Fig. 10), whereas simulated trends ranged from -

410   0.0041% yr^{-1} (INM-CM4-8) to 0.096% yr^{-1} (MPI-ESM1-2-HR), with the MME yielding trend of 0.039 ± 0.032% yr^{-1}. Critically, all models systematically underestimated TCF changes, and none captured the significant increase observed in satellite data.

21
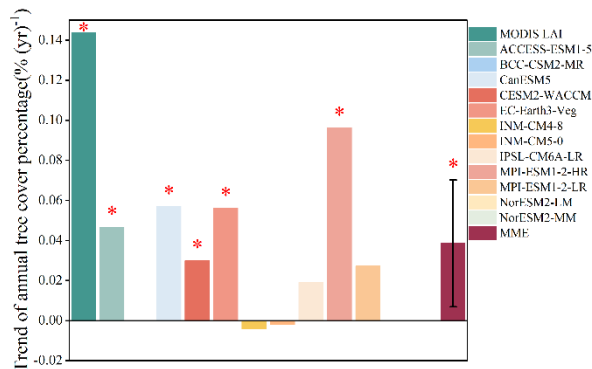
Figure 10: Overall annual average TCF trends in China during 2003~2014. The asterisk (*) indicated the significant trend (p<0.05).

Figure 11 illustrates spatial differences between simulated and observed TCF trends across China. While most models showed reasonable TCF trend simulations in the sparsely vegetated northwestern regions, they overestimated trends on the Tibetan Plateau and underestimated trends in eastern China. The MME underestimated, overestimated, and simulated well TCF trends in 43.72%, 21.47%, and 34.81% in China, respectively. CanESM5 and MPI-ESM1-2-HR demonstrated relatively better simulated TCF trend with observations, well simulating TCF trends in 41.81% and 42.79% of the area, respectively, while overestimating 22.90% and 21.76% and underestimating 35.28% and 35.45% (Fig. 11d and 11i). However, both models still overestimated TCF changes in eastern coastal regions and parts of the Central Plains. Notably, MPI-ESM1-2-LR exhibited the highest proportion of overestimated TCF trends (27.92%) among all models, with 36.24% underestimation. In contrast, INM-CM5-0 showed the most underestimation, underestimating TCF trends in 53.29% of the area while overestimating 12.46%.
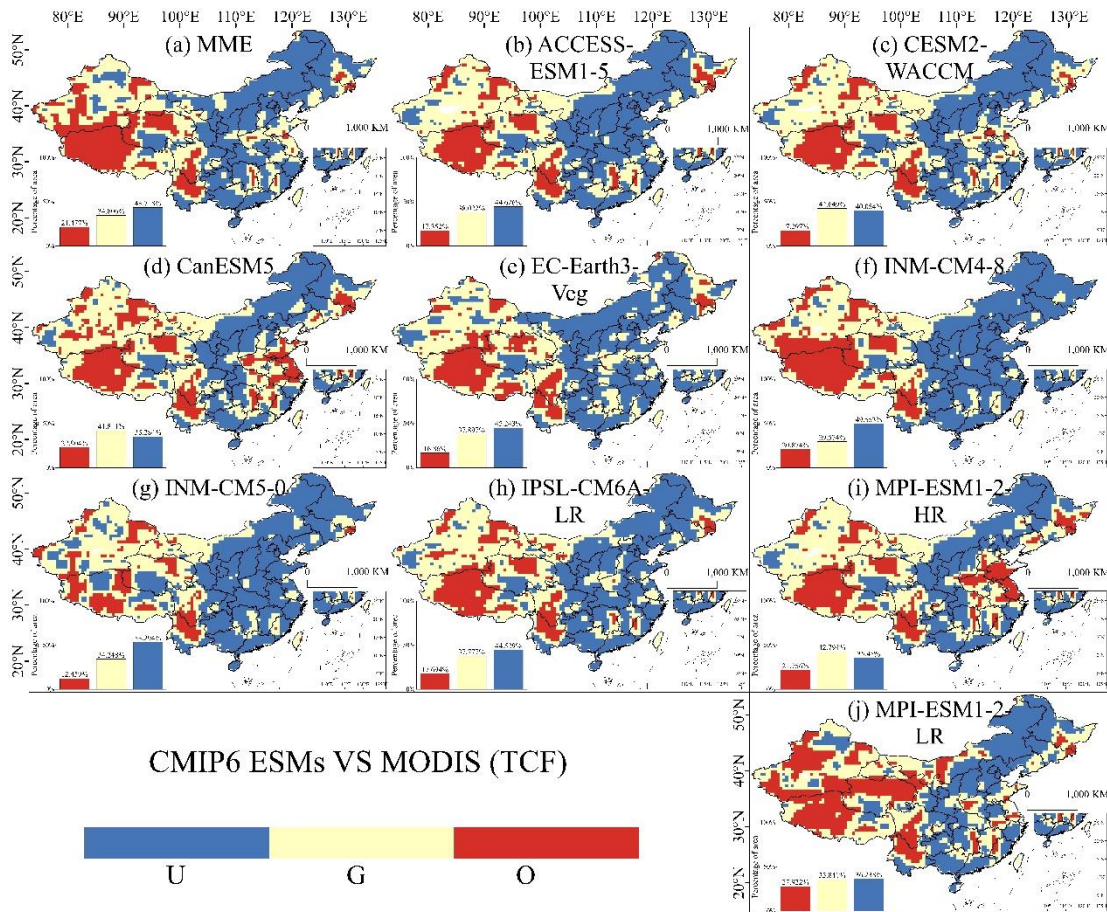
Figure 11: Evaluation of the tree cover prcentage trend performance of the CMIP6 ESM from 2003 to 2014 in China. A paired t-test with a sample size of 9 was conducted using a 3×3 sliding window to determine whether the model simulated a trend that was not significantly different from the observed data (Good), significantly smaller than the observed data (Underestimate), or significantly larger than the observed data (Overestimate). U, G, and O represent 'Underestimate ', 'Good', and 'Overestimate', respectively.

The underestimation in regions with observed significant TCF increases (e.g., afforestation zones in southern China and the Loess Plateau) suggests that CMIP6 models likely fail to adequately represent anthropogenic contributions to vegetation growth, such as afforestation policies and ecological engineering impacts.

4.3 Uncertain of the observed LAI and CSIF in China

Accurate observational data are essential for determining and improving the precision of models (Luo et al., 2016). In this study, two remote sensing datasets that better reflect actual vegetation growth, namely the reprocessed MODIS dataset and the CSIF dataset, were selected for evaluation. However, satellite remote sensing data in China are subject to considerable uncertainty due to various factors. High cloud cover during the rainy season and snow cover in high-latitude areas during winter can introduce inaccuracies. Additionally, satellite sensors are prone to degradation over time, leading to reduced

sensitivity. Although the reprocessed MODIS and CSIF datasets utilize spatiotemporal filtering and machine learning techniques to enhance data quality (Zhang et al., 2018; Yuan et al., 2011), significant uncertainties remain.

5.Conclusion

This study evaluates the trends performance of CMIP6 ESM-simulated LAI, GPP, NPP, NEP and LST in China from 2003
445   to 2014 using MODIS (LAI, NPP, NEP and LST) and CSIF, as observational references. Overall, ESMs fail to capture the trends of CMIP6 ESM-simulated variables in China. In particular, the simulations overestimates and underestimates that are spatially distributed over a large part of the area. The 12 CMIP6 ESMs revealed substantial errors in overall trends in LAI, NPP, and LST, along with significant underestimations in overall trends of GPP and NEP. Discrepancies between observed and simulated trends reached $0.03 \cdot yr^{-1}$ for GPP and 2.46 g $C \cdot m^{-2} \cdot yr^{-1}$ for NEP. Spatially, CMIP6 ESMs underestimated
450   trends in LAI, GPP, NPP, and NEP across China. The MME underestimated these variables in 46.29% (LAI), 43.47% (GPP), 49.81% (NPP), and 61.34% (NEP) of the study area. Concurrently, models predominantly overestimated LST trends in southern China, while predominantly underestimating trends in northern China (MME overestimation: 54.57% and underestimation: 28.02%).

The CMIP6 ESMs exhibit significant misestimation of trends in carbon cycle physical and chemical variables, primarily
455   attributed to inadequate response of environmental factor dynamics and insufficient consideration of anthropogenic influences. This limitation is compounded by the model's failure to adequately represent ecosystem and soil respiration mechanisms, which affects the simulation accuracy of NPP and NEP. Furthermore, systematic misestimation of LAI trend prevents accurate reproduction of the inhibitory effect of LAI on LST.

Code and data availability

460   The R code used for analysis is publicly available. Required packages include rtrend, terra and tidyverse, accessible via CRAN repositories at https://cran.r-project.org/web/packages. All data used in this study are publicly available. The raw CMIP6 data (Eyring et al., 2016) can be downloaded from the USA portal of the Earth System Grid Federation (https://aims2.llnl.gov/search/cmip6). The MODIS LAI data (Yuan et al., 2011) can be downloaded from Data Publisher for Earth & Environmental Science: http://globalchange.bnu.edu.cn/research/laiv061. The CSIF data (Zhang et al., 2018) can be
465   accessed through Figshare: https://doi.org/10.6084/m9.figshare.6387494 (Zhang, 2018). The MODIS NPP data can be download from NASA Land Processes Distributed Active Archive Center: https://doi.org/10.5067/MODIS/MOD17A3H.006 (Running, 2015). The NEP data can be download from Loess Plateau SubCenter, National Earth System Science Data Center, National Science & Technology Infrastructure of China (http://loess.geodata.cn). The LST data (Yu et al., 2022) can be download from National Tibetan Plateau/Third Pole Environment Data Center:
470   https://doi.org/10.11888/Meteoro.tpdc.271663 (Zhao, 2021). The VCF data employed to describe the TCF can be

downloaded from NASA Land Processes Distributed Active Archive Center: https://doi.org/10.5067/MODIS/MOD44B.006 (Dimiceli, 2015).

References

Allen, M., Dube, O. P., Solecki, W., Aragón-Durand, F., Cramer, W., Humphreys, S., and Kainuma, M.: Special report: Global warming of 1.5 C, Intergovernmental Panel on Climate Change (IPCC), 27, 677, 2018.
490 Amthor, J. S.: The McCree–de Wit–Penning de Vries–Thornley Respiration Paradigms: 30 Years Later, Annals of Botany, 86, 1-20, https://doi.org/10.1006/anbo.2000.1175, 2000.
Anav, A., Friedlingstein, P., Kidston, M., Bopp, L., Ciais, P., Cox, P., Jones, C., Jung, M., Myneni, R., and Zhu, Z.: Evaluating the land and ocean components of the global carbon cycle in the CMIP5 earth system models, Journal of Climate, 26, 6801-6843, https://doi.org/10.1175/JCLI-D-12-00417.1, 2013.
495 Aspinwall, M., Jacob, V., Blackman, C., Smith, R., Tjoelker, M., and Tissue, D.: The temperature response of leaf dark respiration in 15 provenances of Eucalyptus grandis grown in ambient and elevated CO2, Functional Plant Biology, 44, 1075-1086, https://doi.org/10.1071/FP17110, 2017.
Bao, Y., Song, Z., and Qiao, F.: FIO-ESM version 2.0: Model description and evaluation, Journal of Geophysical Research: Oceans, 125, e2019JC016036, https://doi.org/10.1029/2019JC016036, 2020.

500   Chen, C., Park, T., Wang, X., Piao, S., Xu, B., Chaturvedi, R. K., Fuchs, R., Brovkin, V., Ciais, P., and Fensholt, R.: China and India lead in greening of the world through land-use management, Nature sustainability, 2, 122-129, https://doi.org/10.1038/s41893-019-0220-7, 2019.

Collalti, A. and Prentice, I. C.: Is NPP proportional to GPP? Waring's hypothesis 20 years on, Tree Physiology, 39, 1473-1483, https://doi.org/10.1093/treephys/tpz034, 2019.

505   DeLUCIA, E. H., DRAKE, J. E., THOMAS, R. B., and GONZALEZ-MELER, M.: Forest carbon use efficiency: is respiration a constant fraction of gross primary production?, Global Change Biology, 13, 1157-1167, https://doi.org/10.1111/j.1365-2486.2007.01365.x, 2007.

Dewar, R. C., Medlyn, B. E., and McMurtrie, R. E.: A mechanistic analysis of light and carbon use efficiencies, Plant, Cell & Environment, 21, 573-588, https://doi.org/10.1046/j.1365-3040.1998.00311.x, 1998.

510   DiMiceli, C., Carroll, M., Sohlberg, R., Kim, D.-H., Kelly, M., & Townshend, J.: MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m SIN Grid V006, NASA Land Processes Distributed Active Archive Center [dataset], https://doi.org/10.5067/MODIS/MOD44B.006, 2015.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development,
515   9, 1937-1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Fu, Z., Ciais, P., Prentice, I. C., Gentine, P., Makowski, D., Bastos, A., Luo, X., Green, J. K., Stoy, P. C., and Yang, H.: Atmospheric dryness reduces photosynthesis along a large range of soil water deficits, Nature communications, 13, 989, https://doi.org/10.1038/s41467-022-28652-7, 2022.

GIFFORD, R. M.: Whole plant respiration and photosynthesis of wheat under increased CO_2 concentration and temperature:
520   long-term vs. short-term distinctions for modelling, Global Change Biology, 1, 385-396, https://doi.org/10.1111/j.1365-2486.1995.tb00037.x, 1995.

Hovenden, M. J., Leuzinger, S., Newton, P. C., Fletcher, A., Fatichi, S., Lüscher, A., Reich, P. B., Andresen, L. C., Beier, C., and Blumenthal, D. M.: Globally consistent influences of seasonal precipitation limit grassland biomass response to elevated CO_2, Nature plants, 5, 167-173, https://doi.org/10.1038/s41477-018-0356-x, 2019.

525   Huang, M., Piao, S., Janssens, I. A., Zhu, Z., Wang, T., Wu, D., Ciais, P., Myneni, R. B., Peaucelle, M., and Peng, S.: Velocity of change in vegetation productivity over northern high latitudes, Nature ecology & evolution, 1, 1649-1654, https://doi.org/10.1038/s41559-017-0328-y, 2017.

Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Camps-Valls, G., Ciais, P., and Friedlingstein, P.: Compensatory water effects link yearly global land CO_2 sink changes to temperature, Nature, 541,
530   516-520, https://doi.org/10.1038/nature20780, 2017.

Kim, D., Lee, M.-I., Jeong, S.-J., Im, J., Cha, D. H., and Lee, S.: Intercomparison of terrestrial carbon fluxes and carbon use efficiency simulated by CMIP5 earth system models, Asia-Pacific Journal of Atmospheric Sciences, 54, 145-163, https://doi.org/10.1007/s13143-017-0066-8, 2018.

Lawrence, D.: Technical description of version 5.0 of the Community Land Model (CLM), (No Title), 329,
535   https://doi.org/10.1029/2018MS001583, 2020.

Le Quéré, C., Raupach, M. R., Canadell, J. G., Marland, G., Bopp, L., Ciais, P., Conway, T. J., Doney, S. C., Feely, R. A., and Foster, P.: Trends in the sources and sinks of carbon dioxide, Nature geoscience, 2, 831-836, https://doi.org/10.1038/ngeo689, 2009.

Lehmann, J. and Kleber, M.: The contentious nature of soil organic matter, Nature, 528, 60-68,
540   https://doi.org/10.1038/nature16069, 2015.

Li, W., Zhang, Y., Shi, X., Zhou, W., Huang, A., Mu, M., Qiu, B., and Ji, J.: Development of land surface model BCC_AVIM2. 0 and its preliminary performance in LS3MIP/CMIP6, Journal of Meteorological Research, 33, 851-869, https://doi.org/10.1007/s13351-019-9016-y, 2019a.

Li, Y., Piao, S., Chen, A., Ciais, P., and Li, L. Z. X.: Local and teleconnected temperature effects of afforestation and
545   vegetation greening in China, National Science Review, 7, 897-912, https://doi.org/10.1093/nsr/nwz132, 2019b.

Li, Y., Zhao, M., Motesharrei, S., Mu, Q., Kalnay, E., and Li, S.: Local cooling and warming effects of forests based on satellite observations, Nature Communications, 6, 6603, https://doi.org/10.1038/ncomms7603, 2015.

Luo, Y., Ahlström, A., Allison, S. D., Batjes, N. H., Brovkin, V., Carvalhais, N., Chappell, A., Ciais, P., Davidson, E. A., Finzi, A., Georgiou, K., Guenet, B., Hararuk, O., Harden, J. W., He, Y., Hopkins, F., Jiang, L., Koven, C., Jackson, R. B.,

Jones, C. D., Lara, M. J., Liang, J., McGuire, A. D., Parton, W., Peng, C., Randerson, J. T., Salazar, A., Sierra, C. A., Smith, M. J., Tian, H., Todd-Brown, K. E. O., Torn, M., van Groenigen, K. J., Wang, Y. P., West, T. O., Wei, Y., Wieder, W. R., Xia, J., Xu, X., Xu, X., and Zhou, T.: Toward more realistic projections of soil carbon dynamics by Earth system models, Global Biogeochemical Cycles, 30, 40-56, https://doi.org/10.1002/2015GB005239, 2016.

Mahowald, N., Lo, F., Zheng, Y., Harrison, L., Funk, C., and Lombardozzi, D.: Leaf area index in Earth system models: evaluation and projections, Earth Syst. Dyn. Discuss, 7, 211-229, https://doi.org/10.5194/esd-7-211-2016, 2016.

Meehl, G. A., Boer, G. J., Covey, C., Latif, M., and Stouffer, R. J.: Intercomparison makes for a better climate model, Eos, Transactions American Geophysical Union, 78, 445-451, https://doi.org/10.1029/97EO00276, 1997.

Nemani, R. R., Keeling, C. D., Hashimoto, H., Jolly, W. M., Piper, S. C., Tucker, C. J., Myneni, R. B., and Running, S. W.: Climate-driven increases in global terrestrial net primary production from 1982 to 1999, science, 300, 1560-1563, https://doi.org/10.1126/science.1082750, 2003.

O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., and Lowe, J.: The scenario model intercomparison project (ScenarioMIP) for CMIP6, Geoscientific Model Development, 9, 3461-3482, https://doi.org/10.5194/gmd-9-3461-2016, 2016.

Park, H. and Jeong, S.: Leaf area index in Earth system models: how the key variable of vegetation seasonality works in climate projections, Environmental Research Letters, 16, 034027, https://doi.org/10.1088/1748-9326/abe2cf, 2021.

Piao, S., He, Y., Wang, X., and Chen, F.: Estimation of China's terrestrial ecosystem carbon sink: Methods, progress and prospects, Science China Earth Sciences, 65, 641-651, https://doi.org/10.1007/s11430-021-9892-6, 2022.

Piao, S., Friedlingstein, P., Ciais, P., Zhou, L., and Chen, A.: Effect of climate and CO2 changes on the greening of the Northern Hemisphere over the past two decades, Geophysical Research Letters, 33, https://doi.org/10.1029/2006GL028205, 2006.

Piao, S., Ciais, P., Friedlingstein, P., de Noblet-Ducoudré, N., Cadule, P., Viovy, N., and Wang, T.: Spatiotemporal patterns of terrestrial carbon cycle during the 20th century, Global Biogeochemical Cycles, 23, https://doi.org/10.1029/2008GB003339, 2009.

Reich, P. B., Hobbie, S. E., and Lee, T. D.: Plant growth enhancement by elevated CO2 eliminated by joint water and nitrogen limitation, Nature Geoscience, 7, 920-924, 2014.

Running, S., Mu, Q., & Zhao, M.: MOD17A3H MODIS/Terra Net Primary Production Yearly L4 Global 500m SIN Grid V006., NASA Land Processes Distributed Active Archive Center [dataset], https://doi.org/10.5067/MODIS/MOD17A3H.006, 2015.

Schuur, E. A. G., Chadwick, O. A., and Matson, P. A.: CARBON CYCLING AND SOIL CARBON STORAGE IN MESIC TO WET HAWAIIAN MONTANE FORESTS, Ecology, 82, 3182-3196, https://doi.org/10.1890/0012-9658(2001)082[3182:CCASCS]2.0.CO;2, 2001.

Slot, M. and Kitajima, K.: General patterns of acclimation of leaf respiration to elevated temperatures across biomes and plant types, Oecologia, 177, 885-900, https://doi.org/10.1007/s00442-014-3159-4, 2015.

Smith, N. G. and Dukes, J. S.: Plant respiration and photosynthesis in global-scale models: incorporating acclimation to temperature and 2, Global Change Biology, 19, 45-63, https://doi.org/10.1111/j.1365-2486.2012.02797.x, 2013.

Song, X., Wang, D.-Y., Li, F., and Zeng, X.-D.: Evaluating the performance of CMIP6 Earth system models in simulating global vegetation structure and distribution, Advances in Climate Change Research, 12, 584-595, 2021.

Sun, X., Xie, W., and Zhou, B.: CMIP6 evaluation and projection of terrestrial ecosystem over Asia, Advances in Climate Change Research, 19, 49, https://doi.org/10.12006/j.issn.1673-1719.2022.045, 2023.

Tang, X., Pei, X., Lei, N., Luo, X., Liu, L., Shi, L., Chen, G., and Liang, J.: Global patterns of soil autotrophic respiration and its relation to climate, soil and vegetation characteristics, Geoderma, 369, 114339, https://doi.org/10.1016/j.geoderma.2020.114339, 2020.

Tian, F., Zhu, Z., Cao, S., Zhao, W., Li, M., and Wu, J.: Satellite-observed increasing coupling between vegetation productivity and greenness in the semiarid Loess Plateau of China is not captured by process-based models, Science of the Total Environment, 906, 167664, https://doi.org/10.1016/j.scitotenv.2023.167664, 2024.

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

Walther, S., Duveiller, G., Jung, M., Guanter, L., Cescatti, A., and Camps-Valls, G.: Satellite observations of the contrasting response of trees and grasses to variations in water availability, Geophysical Research Letters, 46, 1429-1440, https://doi.org/10.1029/2018GL080535, 2019.

Wang, Q., Zheng, H., Zhu, X., and Yu, G.: Primary estimation of Chinese terrestrial carbon sequestration during 2001–2010, Science Bulletin, 60, 577-590, https://doi.org/10.1007/s11434-015-0736-9, 2015.

Wang, X., Piao, S., Ciais, P., Friedlingstein, P., Myneni, R. B., Cox, P., Heimann, M., Miller, J., Peng, S., and Wang, T.: A two-fold increase of carbon cycle sensitivity to tropical temperature variations, Nature, 506, 212-215, https://doi.org/10.1038/nature12915, 2014.

Wu, D., Piao, S., Zhu, D., Wang, X., Ciais, P., Bastos, A., Xu, X., and Xu, W.: Accelerated terrestrial ecosystem carbon turnover and its drivers, Global Change Biology, 26, 5052-5062, https://doi.org/10.1111/gcb.15224, 2020.

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., Jie, W., Zhang, J., Liu, Y., and Zhang, L.: The Beijing climate center climate system model (BCC-CSM): The main progress from CMIP5 to CMIP6, Geoscientific Model Development, 12, 1573-1600, https://doi.org/10.5194/gmd-12-1573-2019, 2019.

Yu, P., Zhao, T., Shi, J., Ran, Y., Jia, L., Ji, D., and Xue, H.: Global spatiotemporally continuous MODIS land surface temperature dataset, Scientific Data, 9, 143, https://doi.org/10.1038/s41597-022-01214-8, 2022.

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 Sensing of Environment, 115, 1171-1187, https://doi.org/10.1016/j.rse.2011.01.001, 2011.

Yuan, W., Xia, J., Song, C., and Wang, Y.-P.: Simulating the land carbon sink: Progresses and challenges of terrestrial ecosystem models, Agricultural and Forest Meteorology, 358, 110264, https://doi.org/10.1016/j.agrformet.2024.110264, 2024.

Zeng, Z., Zhu, Z., Lian, X., Li, L. Z., Chen, A., He, X., and Piao, S.: Responses of land evapotranspiration to Earth's greening in CMIP5 Earth System Models, Environmental Research Letters, 11, 104006, https://doi.org/10.1088/1748-9326/11/10/104006, 2016.

Zhang, H., Zhang, M., Jin, J., Fei, K., Ji, D., Wu, C., Zhu, J., He, J., Chai, Z., and Xie, J.: Description and climate simulation performance of CAS-ESM version 2, Journal of Advances in Modeling Earth Systems, 12, e2020MS002210, https://doi.org/10.1029/2020MS002210, 2020.

Zhang, Y.: CSIF, figshare [dataset], https://doi.org/10.6084/m9.figshare.6387494.v2, 2018.

Zhang, Y., Xu, M., Chen, H., and Adams, J.: Global pattern of NPP to GPP ratio derived from MODIS data: effects of ecosystem type, geographical location and climate, Global Ecology and Biogeography, 18, 280-290, https://doi.org/10.1111/j.1466-8238.2008.00442.x, 2009.

Zhang, Y., Joiner, J., Alemohammad, S. H., Zhou, S., and Gentine, P.: A global spatially contiguous solar-induced fluorescence (CSIF) dataset using neural networks, Biogeosciences, 15, 5779-5800, https://doi.org/10.5194/bg-15-5779-2018, 2018.

Zhao, B., Cao, J., Geng, Y., Zhao, X., and von Gadow, K.: Inconsistent responses of soil respiration and its components to thinning intensity in a Pinus tabuliformis plantation in northern China, Agricultural and Forest Meteorology, 265, 370-380, https://doi.org/10.1016/j.agrformet.2018.11.034, 2019.

Zhao, Q., Zhu, Z., Zeng, H., Zhao, W., and Myneni, R. B.: Future greening of the Earth may not be as large as previously predicted, Agricultural and Forest Meteorology, 292, 108111, https://doi.org/10.1016/j.agrformet.2020.108111, 2020.

Zhao, Q., Zhu, Z., Zeng, H., Myneni, R. B., Zhang, Y., Peñuelas, J., and Piao, S.: Seasonal peak photosynthesis is hindered by late canopy development in northern ecosystems, Nature Plants, 8, 1484-1492, https://doi.org/10.1038/s41477-022-01278-9, 2022.

Zhao, T., Yu, P.: Global daily 0.05 ° spatiotemporal continuous land surface temperature dataset (2002-2022), National Tibetan Plateau Data Center [dataset], https://doi.org/10.11888/Meteoro.tpdc.271663, 2021.

Zhu, Q., Riley, W. J., Tang, J., Collier, N., Hoffman, F. M., Yang, X., and Bisht, G.: Representing nitrogen, phosphorus, and carbon interactions in the E3SM land model: Development and global benchmarking, Journal of Advances in Modeling Earth Systems, 11, 2238-2258, https://doi.org/10.1029/2018MS001571, 2019.

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, Nature Climate Change, 6, 791-795, https://doi.org/10.1038/nclimate3004, 2016.

Ziehn, T., Lenton, A., Law, R. M., Matear, R. J., and Chamberlain, M. A.: The carbon cycle in the Australian Community Climate and Earth System Simulator (ACCESS-ESM1)–Part 2: Historical simulations, Geoscientific Model Development, 650   10, 2591-2614, https://doi.org/10.5194/gmd-10-2567-2017, 2017.