

• Original Paper •

# Assessment of Crop Yield in China Simulated by Thirteen Global Gridded Crop Models

Dezhen YIN<sup>1,2</sup>, Fang LI<sup>1,2</sup>, Yaqiong LU<sup>3</sup>, Xiaodong ZENG<sup>1,2</sup>, Zhongda LIN<sup>4,2</sup>, and Yanqing ZHOU<sup>5,2</sup>

<sup>1</sup>*International Center for Climate and Environment Sciences, Institute of Atmospheric Physics,  
Chinese Academy of Sciences, Beijing 100029, China*

<sup>2</sup>*University of Chinese Academy of Sciences, Beijing 100049, China*

<sup>3</sup>*Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu 610041, China*

<sup>4</sup>*State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics,  
Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China*

<sup>5</sup>*State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography,  
Chinese Academy of Sciences, Urumqi 830011, China*

(Received 25 August 2022; revised 7 May 2023; accepted 7 June 2023)

## ABSTRACT

Global gridded crop models (GGCMs) have been broadly applied to assess the impacts of climate and environmental change and adaptation on agricultural production. China is a major grain producing country, but thus far only a few studies have assessed the performance of GGCMs in China, and these studies mainly focused on the average and interannual variability of national and regional yields. Here, a systematic national- and provincial-scale evaluation of the simulations by 13 GGCMs [12 from the GGCM Intercomparison (GGCMI) project, phase 1, and CLM5-crop] of the yields of four crops (wheat, maize, rice, and soybean) in China during 1980–2009 was carried out through comparison with crop yield statistics collected from the National Bureau of Statistics of China. Results showed that GGCMI models generally underestimate the national yield of rice but overestimate it for the other three crops, while CLM5-crop can reproduce the national yields of wheat, maize, and rice well. Most GGCMs struggle to simulate the spatial patterns of crop yields. In terms of temporal variability, GGCMI models generally fail to capture the observed significant increases, but some can skillfully simulate the interannual variability. Conversely, CLM5-crop can represent the increases in wheat, maize, and rice, but works less well in simulating the interannual variability. At least one model can skillfully reproduce the temporal variability of yields in the top-10 producing provinces in China, albeit with a few exceptions. This study, for the first time, provides a complete picture of GGCM performance in China, which is important for GGCM development and understanding the reliability and uncertainty of national- and provincial-scale crop yield prediction in China.

**Key words:** global gridded crop model, historical crop yield, China, multi-model evaluation

**Citation:** Yin, D. Z., F. Li, Y. Q. Lu, X. D. Zeng, Z. D. Lin, and Y. Q. Zhou, 2024: Assessment of crop yield in China simulated by thirteen global gridded crop models. *Adv. Atmos. Sci.*, **41**(3), 420–434, <https://doi.org/10.1007/s00376-023-2234-3>.

## Article Highlights:

- GGCMs generally underestimate rice yield but overestimate wheat, maize, and soybean yield.
- GGCMs fail to capture the spatial patterns of observed crop yields in China.
- GGCMI models are more skillful in reproducing the interannual variability, while CLM5-crop is better at simulating long-term trends.
- At least one model can skillfully simulate the temporal variability of yield in the top-10 producing provinces, with a few exceptions.

## 1. Introduction

China is the largest producing country of wheat and

rice, the second largest producer of maize, and the fourth largest producer of soybean in the world as of 2020, with a share of about 18%, 28%, 22%, and 6% of the global production, respectively (FAOSTAT, 2022). China has approximately 136 million hectares of cropland, ranking it third in the world, where crop production feeds around 22% of the

\* Corresponding author: Fang LI  
Email: [lifang@mail.iap.ac.cn](mailto:lifang@mail.iap.ac.cn)

world's population (Ghose, 2014; FAO, 2021). With population expansion, economic growth, and urbanization in China, concerns arise regarding its future ability to feed itself and related global impacts (Zhao et al., 2021). Therefore, understanding and predicting the spatiotemporal variability of wheat, maize, rice, and soybean yield in China is vital for food security, global grain trade, and achieving the second goal of UN Sustainable Development.

The temporal and spatial variability of crop yield is affected by climatic and environmental conditions, agronomic management (such as fertilizer application, irrigation, and timing of sowing), and genetic strategies (Kukul and Irmak, 2018). Ray et al. (2015) demonstrated that climate variability explained approximately one-third of yield variability in China. In North China, irrigation (Wang et al., 2008), high levels of soil organic carbon and nitrogen (Tian et al., 2019), and appropriate harvest and sowing dates (Sun et al., 2007) can enhance crop yield. Xiao and Tao (2014) and Liu et al. (2013) disentangled the relative contributions of different factors and highlighted the importance of agronomic managements and cultivar improvement for winter wheat on the North China Plain and rice in East China by combining field experiments with a process-based crop model. Yu et al. (2012) demonstrated using the Agro-C model that rice genetic improvement was the principal driver of China's rice production, especially after nitrogen fertilizer application declined.

Global gridded crop models (GGCMs) are globally consistent and grid- and process-based crop modeling systems that represent crop growth processes and their linear and nonlinear responses to various driving factors in different crop phenological stages (Müller et al., 2017). GGCMs can be separated into two categories: site-based and those used in ecosystem models. Both are derived from a field-scale crop model, but the latter also considers the biophysical and biogeochemical influence of agriculture on the land surface and has become an important component of land surface and Earth system models (Levis et al., 2012; Wu et al., 2016; Müller et al., 2017; Lombardozi et al., 2020). They have been increasingly applied for assessing climate change impacts, adaptation, and environmental impacts of agricultural production (e.g., Lobell et al., 2006; Rosenzweig et al., 2014; Yin et al., 2015; Jägermeyr et al., 2020). Comprehensive evaluation of GGCMs' historical simulations is essential for understanding the reliability of GGCMs in these applications.

The Global Gridded Crop Model Intercomparison (GGCMI) was initiated in 2012, coordinated by the Agricultural Model Intercomparison and Improvement Project and the Inter-Sectoral Impacts Model Intercomparison Project (Elliott et al., 2015). It aims to bring together a diverse international community of crop modelers for climate impact assessment as well as model intercomparison and improvement at the global scale, using the same and standardized protocol (Elliott et al., 2015). So far, the first two phases of GGCMI have been completed, i.e., historical yield simula-

tions (Phase 1, Elliott et al., 2015) and sensitivity simulations of yield responses to different atmospheric CO<sub>2</sub>, temperature, water, and nitrogen levels (Phase 2, Franke et al., 2020). Three simulation scenarios were included in GGCMI phase 1: (1) default: management and technology assumptions and growing seasons that models typically use for simulations in the historical period, presenting the best performance of each model; (2) fullharm: harmonized growing seasons and nitrogen fertilizer inputs; and (3) harmnon: the same as fullharm but with unlimited nitrogen fertilizer supply (Elliott et al., 2015).

Müller et al. (2017) evaluated the average and interannual variability pattern (quantified by temporal correlation) of GGCMI phase 1 default simulations and pointed out that the observed pattern of wheat, soybean, and maize yield can be partly reproduced by these models at the global scale and for top-10 producing countries. Li et al. (2022) assessed the performance of nine models in the GGCMI phase 1 fullharm scenario in simulating the average and interannual variability of yield in seven regions of China and concluded that GGCMI models simulated regional yields of maize better than other grain crops and pointed out the best model of each region. Until now, however, no study has evaluated the performance of GGCMI models in reproducing the observed spatial pattern and trend of crop yield in China (the trend is much stronger than the interannual variability for China's crop yield; see Fig. 3 later in the paper), or on the provincial scale to guide provincial agricultural management.

CLM5-crop is version 5 of the Community Land Model (CLM5) with modeling of actively managed crops (Lawrence et al., 2019; Lombardozi et al., 2020). It was a GGCM developed from CLM4.5post-crop used in GGCMI phase 1 (Levis et al., 2012, 2018) by Lombardozi et al. (2020). It is the only GGCM worldwide used formally in an Earth system model (CESM2) and its code is open to the public. Lombardozi et al. (2020) evaluated its skill on the global scale, but the performance of CLM5-crop in simulating crop yield in China is still unclear.

In this study, we perform a comprehensive national and provincial evaluation of yield simulations of the four main grain crops (i.e., wheat, maize, rice, and soybean) in China made by 12 models from the GGCMI phase 1 default scenario and CLM5-crop to provide a complete picture of GGCM performance. The systematical analyses, which include spatial and temporal evaluations such as the average, spatial pattern, long-term trend, and the magnitude and pattern of interannual variability, are conducted by comparing model simulations with collected national and provincial observations for 1980–2009.

## 2. Data and methods

### 2.1. GGCMs

This study evaluates the performance of 13 GGCMs (Table 1) in simulating crop yield in China. Among them,

**Table 1.** Summary of the GGCMs used in this study.

Crop model	Type <sup>a</sup>	N stress	Crop <sup>b</sup>	Scenarios <sup>c</sup>	Reference(s)
CLM4.5post-crop	E	yes	whe, mai, ric, soy	D, F, H	Levis et al. (2018)
LPJ-GUESS	E	no	whe, mai, ric, soy	D, H	Lindeskog et al. (2013)
LPJmL	E	no	whe, mai, ric, soy	D, H	Bondeau et al. (2007)
ORCHIDEE-crop	E	yes	whe, mai, ric, soy	D, F, H	Wu et al. (2016)
PEGASUS	E	yes	whe, mai, soy	D, F, H	Deryng et al. (2011, 2014)
CGMS-WOFOST	S	no	whe, mai, ric, soy	D	de Wit and van Diepen (2008)
EPIC-Boku	S	yes	whe, mai, ric, soy	D, F, H	Kiniry et al. (1995)
EPIC-IIASA	S	yes	whe, mai, ric, soy	D, F, H	Balkovič et al. (2014)
GEPIC	S	yes	whe, mai, ric, soy	D, F, H	Folberth et al. (2012)
pAPSIM	S	yes	whe, mai, soy	D, F, H	Keating et al. (2003), Elliott et al. (2014)
pDSSAT	S	yes	whe, mai, ric, soy	D, F, H	Jones et al. (2003), Elliott et al. (2014)
PEPIC	S	yes	whe, mai, ric, soy	D, F, H	Liu et al. (2016)
CLM5-crop	E	yes	whe, mai, ric, soy	D	Lombardozi et al. (2020)

<sup>a</sup> E: ecosystem model; S: site-based model

<sup>b</sup> whe: wheat; mai: maize; ric: rice; soy: soybean

<sup>c</sup> D: default; F: fullharm; H: harmnon

12 GGCMs (CLM4.5post-crop, LPJ-GUESS, LPJmL, ORCHIDEE-crop, PEGASUS, CGMS-WOFOST, EPIC-Boku, EPIC-IIASA, GEPIC, pAPSIM, pDSSAT, and PEPIC) are from GGCM phase 1. Simulation data of the 12 GGCM models were obtained from Müller et al. (2019). The 12 models were selected with the exclusion of EPIC-TAMU and PRYSBI2 because EPIC-TAMU did not provide default simulations and PRYSBI2 did not distinguish between rainfed and irrigated crops required by the GGCM protocol (Elliott et al., 2015).

In this study, simulations of the default scenario in GGCM that present the highest simulation skill of each model are used to evaluate the model performance (section 3) and simulations of the fullharm and harmnon scenarios are compared to investigate the simulated response of yield to nitrogen fertilization (section 4). Additionally, simulations of the default scenario are also used to investigate the simulated response of yield to climate in section 4. Furthermore, the GGCM simulations were run at a spatial resolution of 0.5° for 1980–2010 and were driven by the climate dataset AgMERRA (Ruane et al., 2015). Our study uses the simulations for the period 1980–2009 because the yields in 2010 could be missed due to their method of assigning simulations to calendar year (Elliott et al., 2015).

CLM5-crop had several primary developments compared to CLM4.5post-crop used in GGCM phase 1, including the ability to simulate transient crop distribution and crop management due to the introduction of dynamic land units (carbon, nitrogen, water, and energy are conserved during all transitions), phenological triggers that vary with latitude, and tuning of allocation and phenological parameters (Lombardozi et al., 2020). The phenology phases are governed by the growing-degree-day threshold values, and the management strategies include cropland expansion, fertilization, and irrigation (Lombardozi et al., 2020). A detailed description of CLM5-crop can be found in technical documentation

([https://www.cesm.ucar.edu/models/cesm2/land/CLM50\\_Tech\\_Note.pdf](https://www.cesm.ucar.edu/models/cesm2/land/CLM50_Tech_Note.pdf)).

In this study, we performed a transient 1850–2009 run of CLM5-crop (component name: IHistClim50BgcCrop). The simulations were driven by atmospheric forcing (precipitation, temperature, specific humidity, surface pressure, wind speed, and solar radiation) from GSWP3v1, with a 0.9° (latitude) × 1.25° (longitude) horizontal resolution. Crop distribution, industrial fertilizer, and irrigated area were derived from Land Use Harmonization Version 2 (Hurt et al., 2011).

## 2.2. Benchmarks

The wheat, maize, rice, and soybean yields of 31 provinces (autonomous regions or municipalities) during 1980–2009 were collected from the National Bureau of Statistics of China (NBSC, <https://data.stats.gov.cn/easyquery.htm?cn=E0103>). The planting area and nitrogen fertilizer application rate were also obtained from the NBSC. The growing season of wheat, maize, and rice obtained from ChinaCropPhen1km (Luo et al., 2020) and climate data obtained from CN05.1 (Wu and Gao, 2013) were used to calculate growing season temperature and precipitation.

## 2.3. Aggregating simulated yield data

In order to compare with the yield observations, which were available at provincial and national scales, gridded simulations were aggregated into provincial and national yield  $Y(t)$  for year  $t$  for each grain crop type using the area-weighted average method:

$$Y(t) = \frac{\sum_{i=1}^n [a_{i,rf}(t)y_{i,rf}(t) + a_{i,ir}(t)y_{i,ir}(t)]}{\sum_{i=1}^n [a_{i,rf}(t) + a_{i,ir}(t)]}. \quad (1)$$

Here,  $y$  is the gridded simulated yield and  $a$  is the harvested area in grid cell  $i$  under fully irrigated (ir) or purely rainfed (rf) conditions for GGCM models and for irrigated crop functional types (CFTs) and rainfed CFTs for CLM5-crop. Because the GGCM models simulated crops on all agricultural lands no matter which type of crop was planted and whether irrigated or rainfed methods were used in the real world (Elliott et al., 2015), MIRCA2000, with a spatial resolution of 30 arc minutes (Portmann et al., 2010), was used to calculate the harvested area of gridded irrigated/rainfed crop types, following Müller et al. (2017).

#### 2.4. Statistical analyses

This study performs a systematic evaluation of GGCMs in simulating crop yield in China, including the multi-year average of national yield, spatial pattern, long-term trend, pattern and magnitude of interannual variability, and temporal variability of yield in top-10 producing provinces.

The Pearson correlation coefficient is used to evaluate the ability (skill) of models to reproduce the observed spatial or temporal variability pattern, following many earlier evaluation studies (e.g., Sperber et al., 2013; Müller et al., 2017; Li et al., 2019). The Pearson correlation coefficient is a number between  $-1$  and  $1$  that measures the strength and direction of the linear relationship between two variables. The simulation skill is higher if the Pearson correlation coefficient between simulations and observations is closer to  $1$ . Student's  $t$ -test was used to assess its significance. If the Pearson correlation coefficient between simulations and observations is positive and significant, then the simulations are skillful. Spatial correlation is used to evaluate the similarity of spatial distribution between observations and simulations, while temporal correlation is used to evaluate the similarity of observed and simulated time series for a province or for the national average.

We estimated the long-term trend using the ordinary least-squares method. The significance was tested using the Mann–Kendall test.

The detrended component of time series was regarded as the interannual variability. Its amplitude and pattern were quantified using the coefficient of variation (CV, the standard deviation divided by the mean) and the Pearson correlation coefficient, respectively.

As the multi-model ensemble (MME) mean can represent the overall performance of multiple models and often performs better in reproducing observations than individual models (Martre et al., 2015; Wallach et al., 2018), the MME was calculated as the average of the GGCMs. Furthermore, we used standard deviation (SD) to estimate the intermodel spread.

The normalized root-mean-square error (RMSE) [Eq. (2); Li et al., 2022] of anomalies was applied to evaluate the temporal variability of simulated yield for top-10 producing provinces. The mean null model (temporal mean of observations) (Hantson et al., 2020), also called the climatology forecast in meteorology (Li, 2011), was used to assess whether a GGCM was skillful.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n [(y_{\text{sim},i} - \bar{y}_{\text{sim}}) - (y_{\text{obs},i} - \bar{y}_{\text{obs}})]^2}{n}}, \quad (2)$$

where  $y_{\text{sim},i}$  and  $y_{\text{obs},i}$  are the simulated and observed yield in year  $i$ , and  $n$  is the number of years.

### 3. Results

#### 3.1. Multi-year average of national yield

GGCM models generally underestimate national rice yield but overestimate yield for the other three crops [Fig. 1, Table S1, in the Electronic Supplementary Material (ESM)]. The observed, 1980–2009 averaged national crop yields are  $3.51 \times 10^{-1}$ ,  $4.49 \times 10^{-1}$ ,  $5.79 \times 10^{-1}$ ,  $1.53 \times 10^{-1}$  kg m<sup>-2</sup> for wheat, maize, rice, and soybean, respectively, while the GGCM MME yields are  $3.94 \times 10^{-1}$ ,  $5.59 \times 10^{-1}$ ,  $4.36 \times 10^{-1}$ ,  $2.16 \times 10^{-1}$  kg m<sup>-2</sup> (Table S1, in the ESM). The MMEs overestimate the crop yield of wheat, maize, and soybean by 12%, 24%, and 41%, respectively, but underestimate rice by 25%. Among the twelve GGCM models, seven, nine, and ten models overestimate the yield of wheat, maize, and soybean, among which CGMS-WOFOST simulates yields more than twice those observed for soybean (Figs. 1a, b and d). Seven out of ten models underestimate the rice yield, among which CLM4.5post-crop, LPJ-GUESS, and EPIC-Boku simulate a yield less than half that observed (Fig. 1c). EPIC-IIASA performs the best for wheat, maize, and rice, and PEGASUS does the best for soybean (Fig. 1). The GGCM models show large intermodel spread, with the maximum spread of  $2.26 \times 10^{-1}$  kg m<sup>-2</sup> for rice and the minimum of  $0.84 \times 10^{-1}$  kg m<sup>-2</sup> for soybean (Table S1, in the ESM).

CLM5-crop reproduces the national yields of wheat, maize, and rice well, but overestimates the soybean yield by 45% (Fig. 1). It outperforms most GGCM models for wheat, maize, and rice, and is close to or even better than the GGCM-MME (Fig. 1, Table S1, in the ESM).

#### 3.2. Spatial pattern

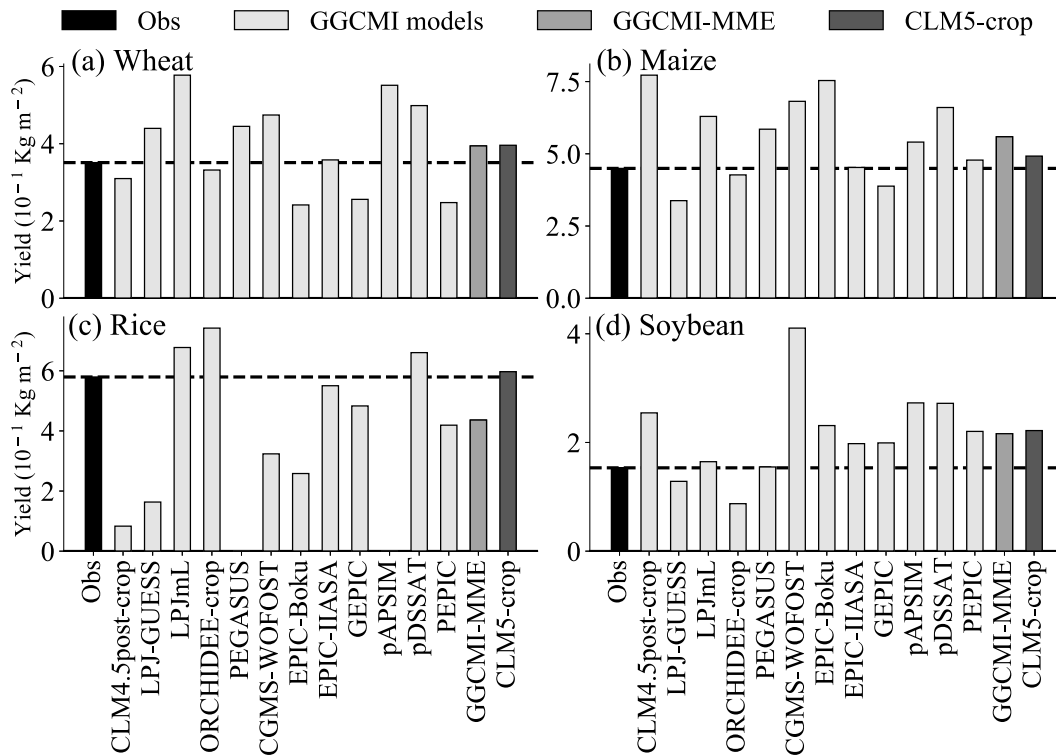
Most of the GGCMs fail to reproduce the spatial patterns of crop yields (Fig. 2). No model can skillfully simulate the spatial pattern for wheat. Only one model is skillful for maize (PEGASUS) and for rice (EPIC-IIASA), and two models are skillful for soybean (LPJmL and CLM5-crop) (Fig. 2). Some of the models even simulate a significant negative correlation (e.g., EPIC-Boku for wheat, LPJmL for maize) (Figs. S1h and S2d, in the ESM). Most of the GGCMs overestimate the yield in Southwest China for wheat (Fig. S1, in the ESM) and maize (Fig. S2, in the ESM). The models with significant negative spatial correlation for rice (CLM4.5post-crop and LPJ-GUESS) tend to simulate high yields in southern China (Figs. S3b and S3c, in the ESM).

#### 3.3. Temporal variability

##### 3.3.1. Trend

The observed national yields show significant increases





**Fig. 1.** Multi-year average national yield (units:  $10^{-1} \text{ kg m}^{-2}$ ) for (a) wheat, (b) maize, (c) rice, and (d) soybean, from observations and GGCM simulations for 1980–2009. See Table S1 (in the ESM) for values corresponding to the bars.

during 1980–2009 for all four crop types (Fig. 3), with trends of  $0.080 \times 10^{-1}$ ,  $0.074 \times 10^{-1}$ ,  $0.065 \times 10^{-1}$ , and  $0.020 \times 10^{-1} \text{ kg m}^{-2} \text{ yr}^{-1}$  for wheat, maize, rice, and soybean, respectively (Table S2 in the ESM). The long-term trend is much stronger than the interannual variability for wheat, maize, and rice (Figs. 3a–c), with the SDs of the former being 3.6, 2.2, and 2.3 times the latter, respectively.

Most of the GGCM models and the MME fail to capture the significant upward trends (Figs. 3–4). Some of the ecosystem models, as well as CGMS-WOFOST, can capture the significant increases, but with lower trend magnitude, except CLM4.5post-crop and PEGASUS for soybean (Fig. 4). Contrary to the observations, all of the GGCM models except PEGASUS simulate a decreasing trend in national maize yield (Fig. 4b).

CLM5-crop simulates an upward trend in national yield for all crop types and successfully captures the significant increases in wheat, maize, and rice, even though it also underestimates the trend magnitude for wheat and maize (Fig. 4). It outperforms 11 of the 12 GGCM models for wheat, maize, and all models for rice, as well as the GGCM-MME for wheat, maize, and rice (Fig. 4, and Table S2 in the ESM).

Similar to the above national results, observed yields exhibit a significant increasing trend over most provinces for the four crop types (Figs. 5a–d). The GGCM MME fails to capture the significant upward trend in most provinces (Figs. 5e–h), whereas individually, CLM4.5post-crop does (for wheat) and CLM4.5post-crop and PEGASUS

also do (for soybean) (Figs. S5a, S8a, and S8e, in the ESM). In contrast, CLM5-crop captures the significant upward trend in most provinces for wheat, maize, and rice, but generally underestimates the trend magnitude, except for rice, and incorrectly locates the high values for wheat, rice, and soybean (Figs. 5i–l).

### 3.3.2. Interannual variability

The magnitude of the interannual variability of the observed yield is weak, with CVs of 0.06 for wheat, 0.07 for maize, 0.04 for rice, and 0.09 for soybean (Table S3, in the ESM). Overall, the GGCMs can reproduce the small values, except for PEGASUS (CV = 0.11), CGMS-WOFOST (CV = 0.18), and pDSSAT (CV = 0.23) for wheat, and CLM4.5post-crop (CV = 0.17) for rice (Fig. 6, Table S3 in the ESM). Several models reproduce less than half the observed interannual variation, e.g., pAPSIM (CV = 0.02) for wheat, CLM4.5post-crop (CV = 0.03) and ORCHIDEE-crop (CV = 0.03) for maize, ORCHIDEE-crop (CV = 0.01) for rice, and PEPIC (CV = 0.04) for soybean (Table S3, in the ESM). CLM5-crop performs better than GGCM MME for maize, rice, and soybean (Fig. 6).

Spatially, the observational interannual variability is higher in the northeastern region of China for wheat and maize, northeastern and southwestern regions for rice, and middle reaches of the Huanghe River, northeastern region, and southwestern region for soybean (Fig. S9, in the ESM). The GGCM MME simulates the spatial pattern of CV for maize (Fig. S9f, in the ESM) and soybean (Fig. S9h, in the

	Wheat	Maize	Rice	Soybean	
CLM4.5post-crop	-0.52	0.16	-0.57	0.29	
LPJ-GUESS	0.51	-0.15	-0.54	0.21	
LPJmL	0.42	-0.57	0.11	0.57	
ORCHIDEE-crop	0.44	0.41	0.06	0.38	
PEGASUS	-0.25	0.51		-0.3	
CGMS-WOFOST	0.43	0.18	0.48	-0.25	Sig. +
EPIC-Boku	-0.67	-0.25	0.03	0.4	
EPIC-IIASA	-0.17	0.26	0.58	0.23	+
GEPIC	0.16	0.3	0.01	0.29	-
pAPSIM	0.53	0.35		-0.1	
pDSSAT	0.06	0.35	-0.23	0.17	
PEPIC	-0.4	-0.07	0.39	-0.03	
GGCMI-MME	0.04	0.31	0.28	0.29	
CLM5-crop	-0.13	-0.35	0.23	0.6	

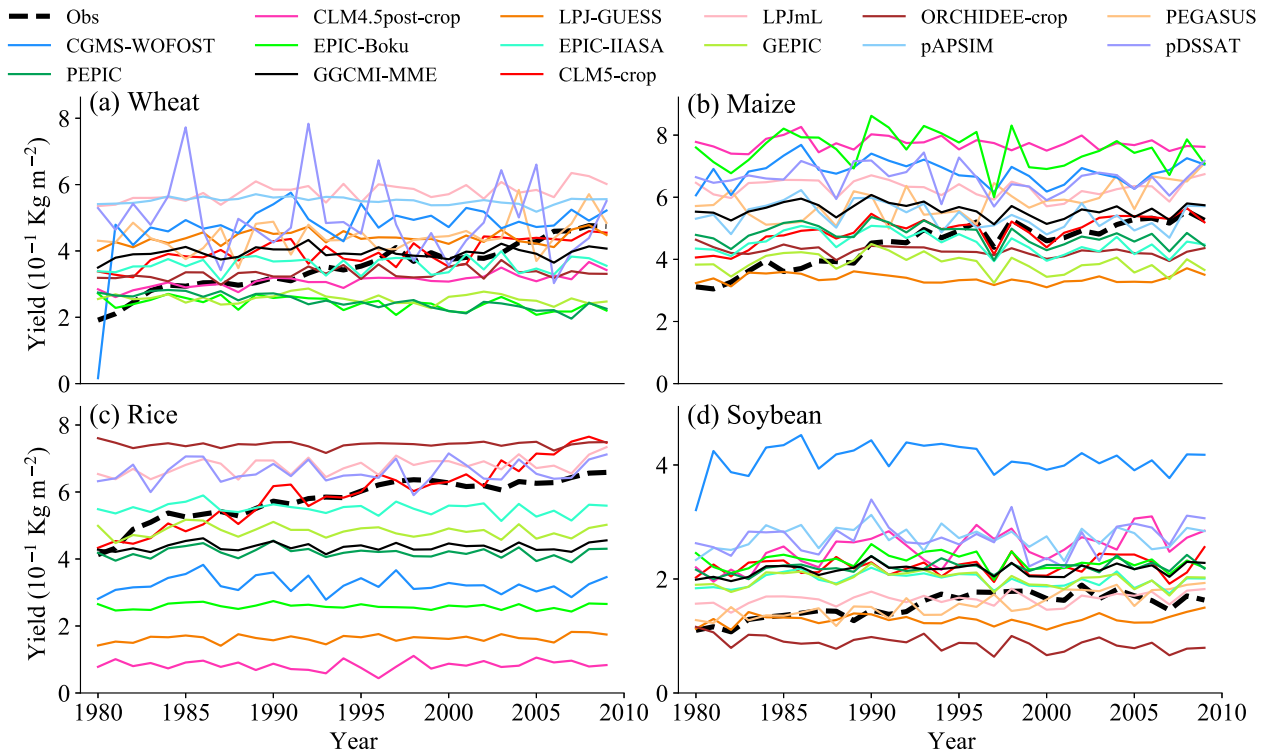
**Fig. 2.** Spatial correlation between observations and simulations for four crop types. Only provinces where production accounts for more than 1% of the national total are considered. See Figs. S1–S4 (in the ESM) for spatial distribution of crop yields corresponding to this figure.

ESM) reasonably well. The GGCMI models simulate the spatial pattern best for maize, with seven out of twelve models showing significant positive spatial correlation with observations (Fig. S11, in the ESM). Furthermore, those models with significant positive spatial correlation can generally capture the high CVs in the northeastern region, while those models with significant negative spatial correlation are inclined to simulate high CVs in the southern region (Figs. S10–S13, in the ESM). CLM5-crop can reproduce the spatial pattern of CV for rice and soybean reasonably well, with a significant spatial correlation of 0.55 and 0.61, respectively (Figs. S9k–l, in the ESM). For wheat (Fig. S9i, in the ESM) and maize (Fig. S9j, in the ESM), CLM5-crop generally simulates relatively high CVs in northwestern China, where the observational CVs are actually low.

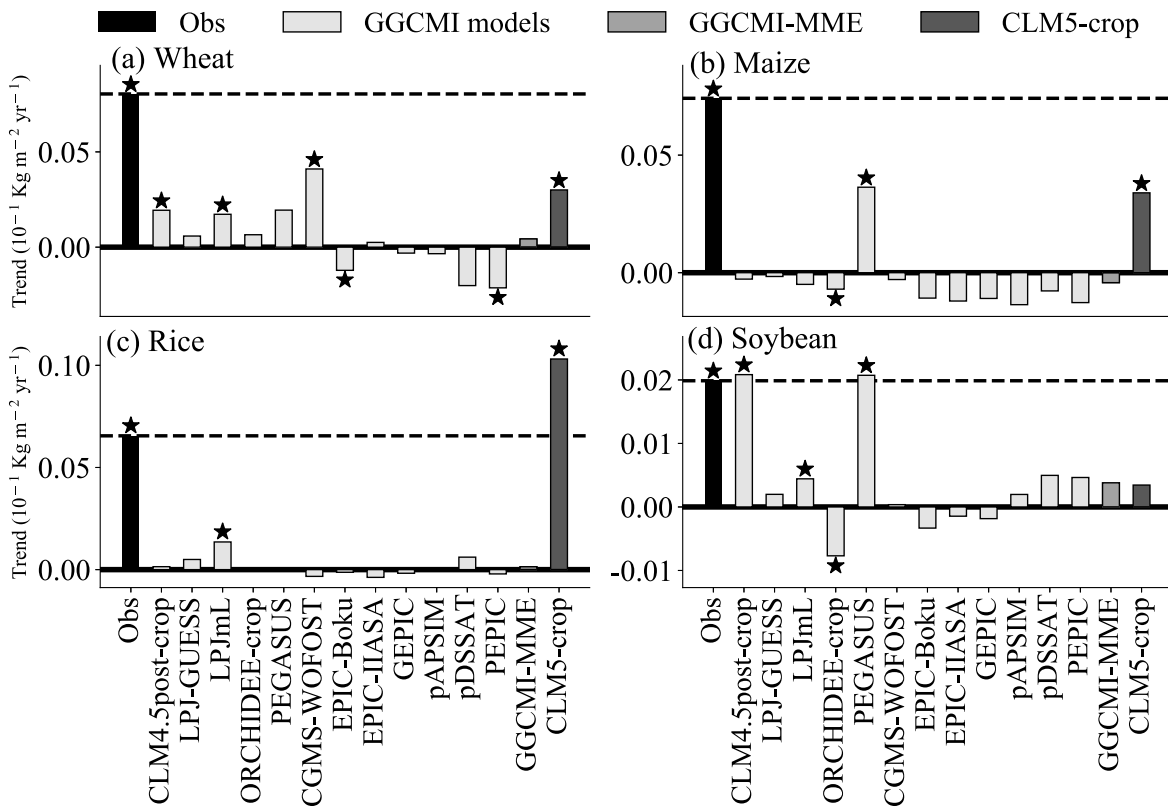
GGCMI site-based models generally perform better in terms of modeling the interannual variability pattern of national yield, especially for maize (Fig. 7). For GGCMI models, all significant positive correlations are simulated by site-based models (Fig. 7). The MME of GGCMI site-based

models shows higher simulation skill for all four crops (0.23, 0.57, 0.33, and 0.34) than the MME of GGCMI ecosystem models (0.1, -0.02, 0.05, and -0.06) (Fig. 7, Table S4, in the ESM). The site-based models perform the best for maize, with four out of seven models exhibiting significant positive correlation with observations and a significant correlation of 0.57 for the site-based MME (Fig. 7, Table S4, in the ESM). CLM5-crop can simulate the interannual variability pattern of maize well, with a significant correlation of 0.64, but its skill for the other three crops is poor, with insignificant temporal correlations of 0.16, 0.18, and 0.09 for wheat, rice, and soybean (Table S4, in the ESM).

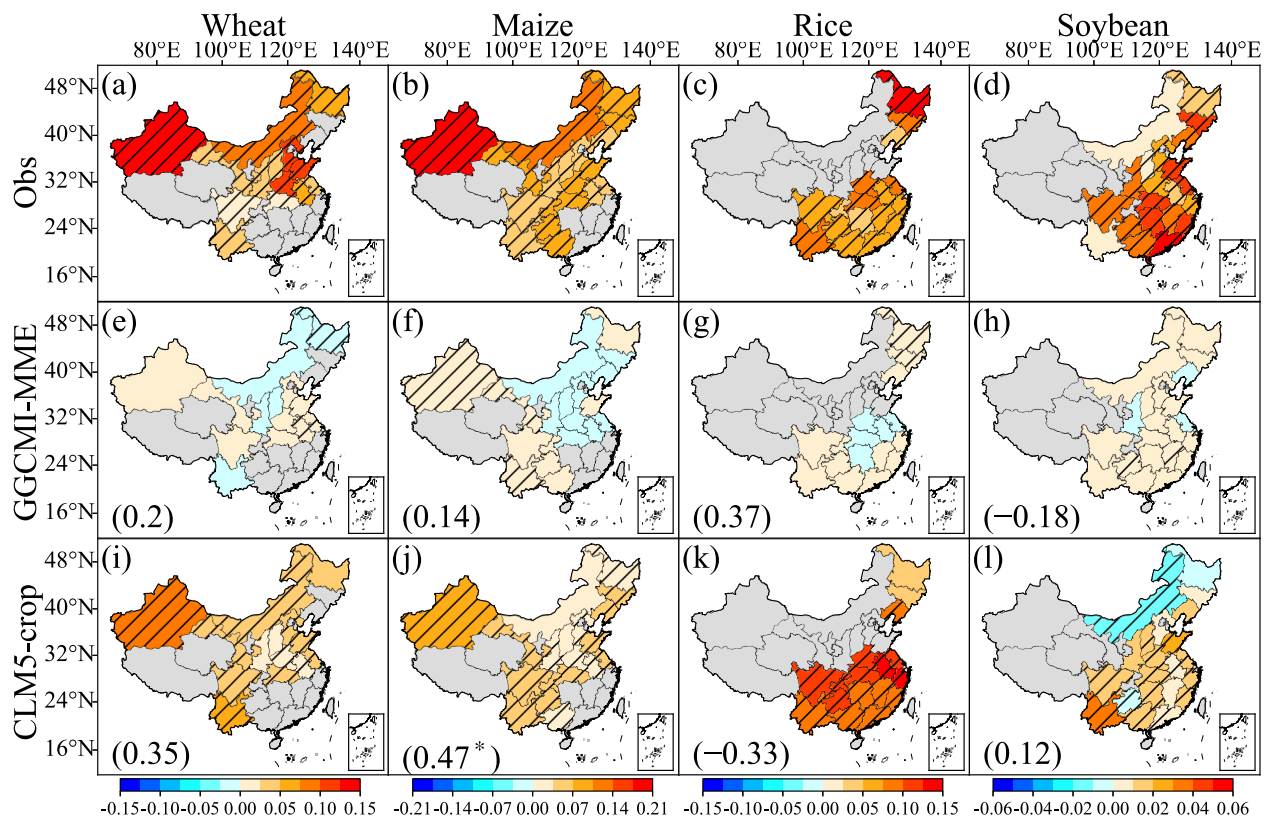
In terms of provincial yield, GGCMI ecosystem-model MME is unable to skillfully simulate the interannual variability pattern, especially for wheat and rice (Fig. 8). GGCMI site-based MME, meanwhile, simulates the interannual variability of provincial maize and soybean yield better, with average correlations more than twice those of wheat and rice (Figs. 8e–h). Provinces with simulation skill for wheat, maize, and soybean are mainly located in North China for



**Fig. 3.** Temporal change in observed and simulated yield for (a) wheat, (b) maize, (c) rice, and (d) soybean in China in 1980–2009. Lines in warm colors are for ecosystem models and lines in cool colors are for site-based models.



**Fig. 4.** Long-term trend (units:  $10^{-1} \text{ kg m}^{-2} \text{ yr}^{-1}$ ) of observed and simulated national yield for (a) wheat, (b) maize, (c) rice, and (d) soybean in 1980–2009. A star above the bar denotes statistical significance according to the Mann–Kendall test at the 0.05 level. See Table S2 (in the ESM) for values corresponding to the bars.



**Fig. 5.** Spatial distribution of the long-term trend (units:  $10^{-1} \text{ kg m}^{-2} \text{ yr}^{-1}$ ) of observations, the MME of GGCM models, and CLM5-crop for four crops during 1980–2009. Provinces with a significant trend ( $p < 0.05$ ) are striped. Provinces where production accounts for more than 1% of the national total are considered and the numbers in parentheses are the spatial correlation coefficient between observed and simulated trends. An asterisk (\*) denotes statistical significance according to Student's  $t$ -test at the 0.05 level. See Figs. S5–S8 (in the ESM) for the spatial distribution of individual GGCM models.

the site-based models (Figs. S14, S15, and S17, in the ESM). Moreover, CLM5-crop can skillfully simulate the interannual variability of provincial maize yield in North China (Fig. 8j).

### 3.3.3. Top-10 producing provinces

We used the RMSE of yield anomalies to test the skill of models in simulating the temporal variability of yield for top-10 producing provinces, to serve as model selection for crop yield forecasting/projection and agricultural management of the major crop-producing provinces. As shown in Fig. 9, there is at least one model whose simulation is skillful for the top-10 producing provinces, except for wheat in Shaanxi and soybean in Heilongjiang and Hebei, though most of the GGCMs cannot simulate the interannual variability skillfully (i.e., worse than the mean null model, depicted by the white boxes). The best models for wheat, maize, and rice are mostly the CLM family of models (CLM5-crop or CLM4.5post-crop) (Figs. 9a–c). The best models for soybean are PEPIC, GGCM-MME, PEGASUS, CLM5-crop, and LPJmL (Fig. 9d). The skill is improved from CLM4.5post-crop to CLM5-crop mainly for maize and rice.

For wheat, LPJmL clearly produces the best yield simulations for Henan, Shandong, and Anhui, while the CLM family of models works best for the other seven provinces (Fig. 9a),

mainly because they can reproduce the significant upward trends (Fig. 5i, Fig. S5a, and Fig. S5c, in the ESM). For maize, CLM4.5post-crop and PEPIC perform the best for Liaoning and Shanxi, respectively, because they are skillful in simulating the interannual variability pattern (Fig. 9b, Fig. S15a, and Fig. S15l, in the ESM). CLM5-crop shows the best simulations for the remaining eight provinces, mainly because it successfully captures the significant increases in yield in these provinces, except for Inner Mongolia, and is skillful in simulating the interannual variability pattern for Jilin, Shandong, Hebei, and Inner Mongolia (Fig. 5j, Fig. 8j and Fig. 9b). For rice, LPJ-GUESS performs the best for Hunan, and LPJmL for Anhui and Zhejiang, mainly because they simulate the interannual variability pattern of rice yield well in these provinces (Fig. 9c and Figs. S16b–c, in the ESM). GEPIC and EPIC-IIASA have the best simulations for Sichuan and Heilongjiang, respectively, while CLM5-crop does for the remaining five provinces, mainly because they successfully capture the significant upward trends (Fig. 5k, Fig. 9c, and Figs. S7g–h, in the ESM). For soybean, LPJmL works the best for Hebei, and GGCM-MME for Henan, Inner Mongolia, and Sichuan, mainly due to the well reproduced patterns of interannual variability (Fig. 8h, Fig. 9d and Fig. S17c, in the ESM). CLM5-crop is the most skillful for Shandong, mainly due to the simulated significant



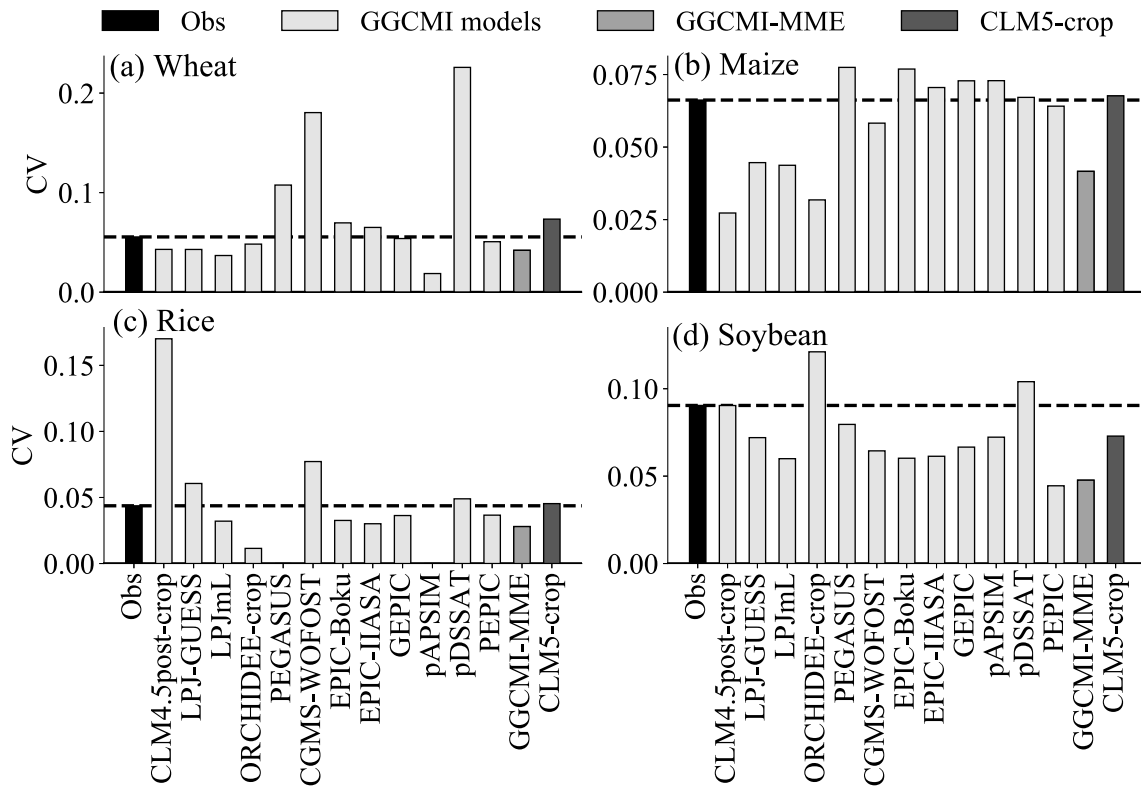


Fig. 6. As in Fig. 1 but for the amplitude of interannual variability, which is calculated using the coefficient of variation (CV) of detrended time series. See Table S3 (in the ESM) for values corresponding to the bars, and Figs. S9–S13 (in the ESM) for spatial distribution of CVs.

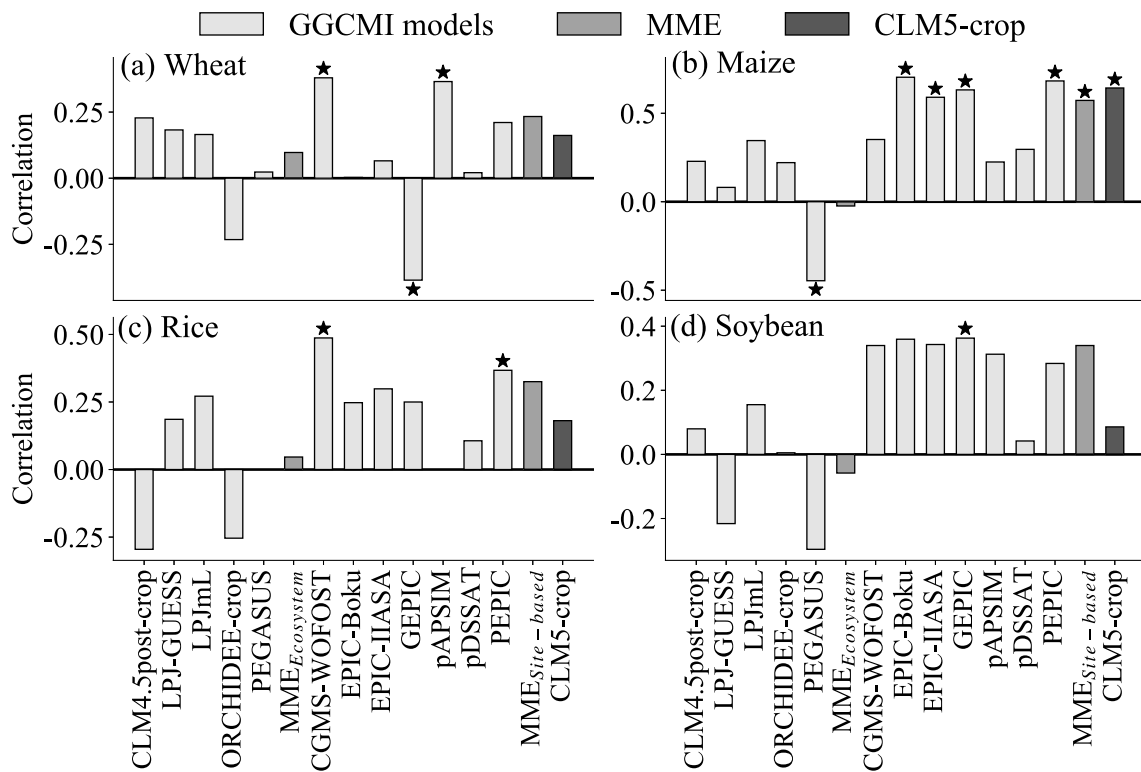
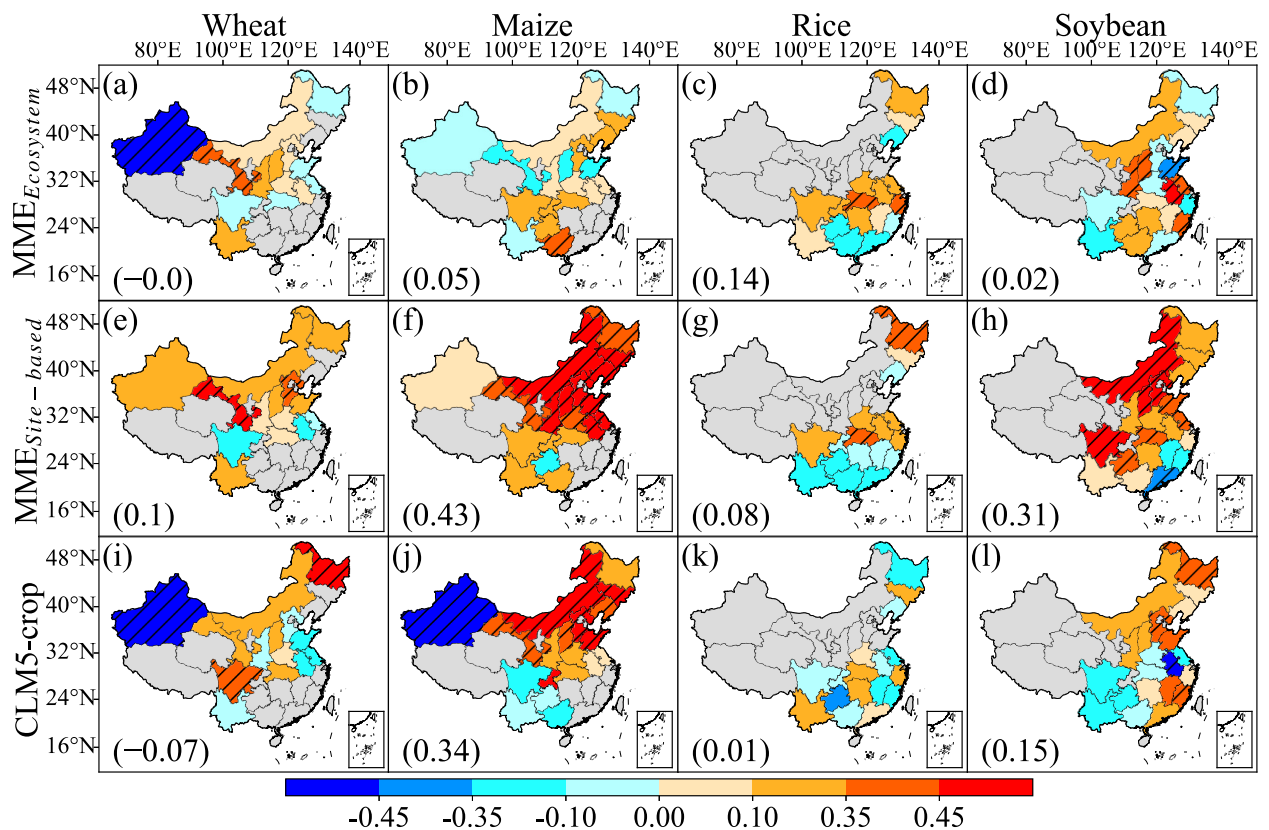


Fig. 7. As in Fig. 4 but for the temporal correlation of the interannual variability of national yield between observations and simulations based on detrended time series. See Table S4 (in the ESM) for values corresponding to the bars. MME<sub>Ecosystem</sub> and MME<sub>Site-based</sub> are the MME means of the GGCMI ecosystem models and site-based models, respectively.



**Fig. 8.** As in Fig. 5 but for the spatial distribution of the temporal correlation coefficient between observed and simulated yields during 1980–2009. Numbers in parentheses are the weighted average of the provincial temporal coefficient with planting area as the weight. See Figs. S14–S17 (in the ESM) for the spatial distribution of individual GGCM models.

upward trend (Fig. 5l and Fig. 9d). PEPIC shows the best yield simulations for Northeast China, mainly because it successfully captures the significant increases in Heilongjiang and Jilin and simulates the interannual variability pattern well in Liaoning (Fig. 9d, Fig. S8l, and Fig. S17l, in the ESM). PEGASUS shows the best simulations for Anhui and Jiangsu, due to its good performance in simulating not only the significant upward trend but also the pattern of interannual variability (Fig. 9d, Fig. S8e, and Fig. S17e, in the ESM).

#### 4. Discussion

The finding that GGCM models fail to capture the significant upward trends in wheat, maize, and rice yield is partly because of the incorrect input data of nitrogen fertilizer supply and/or inaccurate simulations of yield response to agronomic management and cultivar improvement. Nitrogen is an important nutrient required for crop growth, and nitrogen fertilization can alleviate the nitrogen limitation in photosynthesis and nitrogen competition between microorganisms and crops (Sinclair and Rufty, 2012; Lombardozzi et al., 2020). As shown in Fig. S18 (in the ESM), GGCM input data do not consider the observed significant increase in nitrogen fertilizer supply in China for 1980–2009, which leads to the underestimated trends in the GGCM models that consider nitrogen stress (e.g., CLM4.5post-crop, ORCHIDEE-crop,

PEGASUS, site-based GGCMs except for CGMS-WOFOST). Furthermore, many of the GGCM models often simulated lower yields in the harmnon scenario (no nitrogen stress) than in the fullharm scenario in China (Figs. S19–S21, in the ESM), suggesting inaccurate simulations of the yield response to increased nitrogen fertilization, possibly due to inaccurate assumptions of the fertilization scheme (e.g., timing, duration, and rate) and biogeochemical scheme (e.g., photosynthesis and nitrogen cycle). Besides, LPJ-GUESS, LPJmL, and CGMS-WOFOST in GGCM Phase 1 did not model nitrogen stress. On the contrary, CLM5-crop uses a nitrogen fertilizer input that rises significantly as observed (Fig. S18, in the ESM), and models the nitrogen cycle and nitrogen–carbon interactions, which enables it to capture the significant upward yield trends of the three crop types. In addition, no GGCM considers the improvement in management level (e.g., developments in mechanization, inter-cropping, multiple-growing season, intra-annual rotation, or crop residue management) or the changes in crop cultivars and policies, which are important for yield increases in China over the past few decades (Yu et al., 2012; Liu et al., 2013; Xiao and Tao, 2014).

Many GGCMs fail to reproduce the spatial pattern and interannual variability of crop yield in China. This may be partly caused by their poor performance in capturing the response of crop yield to climate conditions. For example,

### RMSE

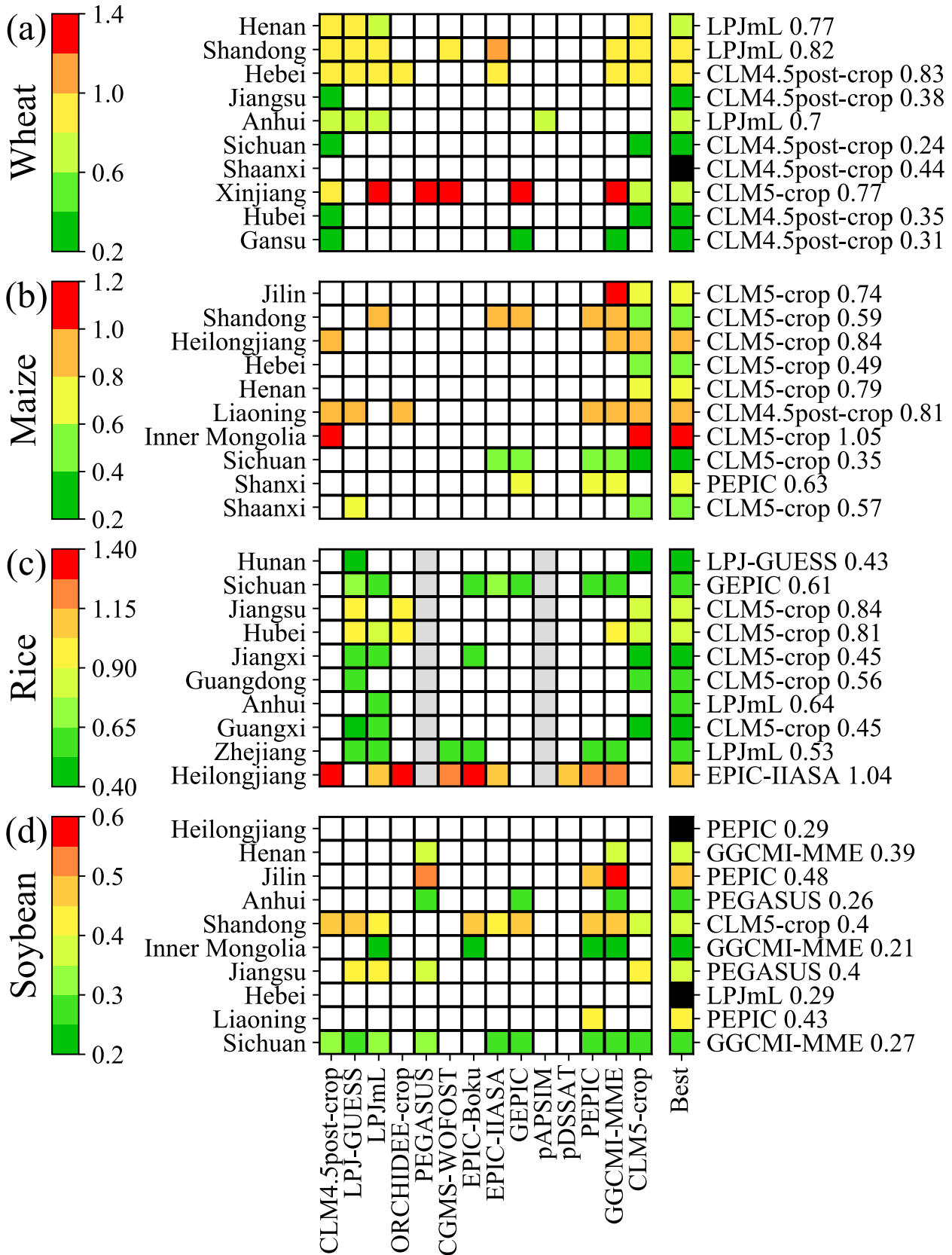


Fig. 9. RMSEs of the time series of yield anomalies simulated by GGCs for top-10 producing provinces in China. Right-hand labels describe the best performing GGCs for that province and the RMSE. White boxes indicate the skill is worse than the mean null model (temporal mean of observations, also called the climatology forecast in meteorology).

CGMS-WOFOST, EPIC-IIASA, and PEPIC can successfully reproduce the observed yield–climate relationship for rice, while other models perform poorly (Fig. S22, in the ESM). Correspondingly, the simulation skill of these three models is much better than others in simulating the spatial pattern and interannual variability of rice yield, with their spatial and temporal correlations being higher than 0.3 (Fig. 2, and Table S4 in the ESM). Similar conclusions can be made for wheat and maize (Figs. S23 and S24, in the ESM). In addition, GGCMs make a number of simplifications in terms of sowing date, cultivars, and response to extreme weather disasters (Barlow et al., 2015; Müller et al., 2017; Rötter et al., 2018; Heinicke et al., 2022), which may partly explain their poor capability in reproducing the spatial pattern and interannual variability of crop yield.

Earlier studies (Müller et al., 2017; Li et al., 2022) that evaluated GGCM models in simulating the average and interannual variability patterns of national and regional yields in China showed that models generally underestimate rice yield but overestimate the yields for the other three crop types in China, which is consistent with our results. They also pointed out that GGCM models perform better for maize than other crop types in their simulation of the interannual variability pattern, and our results also show that GGCM site-based models perform the best for maize, with four out of seven site-based models exhibiting significant positive temporal correlation and a significant temporal correlation of 0.57 for the site-based MME (Fig. 7 and Table S4, in the ESM). However, our study also evaluated the spatial patterns and long-term trends of crop yields, thereby providing a complete picture of model performance, which is critical for comprehensively estimating the capability of models in predicting and projecting yields. Besides, our study, for the first time, evaluated the temporal change in yield for top-10 producing provinces, and CLM5-crop in simulating China's crop yield. In China, provincial agricultural management is largely independent. Therefore, information on model skill and the best performing models at the provincial scale can serve as a guide for model platform selection in predicting and projecting future provincial crop production and carrying out numerical experiments for adjusting and developing agronomic measures. CLM5-crop is the updated version of CLM4.5post-crop in the GGCM project, and the only one formally used in an Earth system model. Lombardozi et al. (2020) evaluated the performance of CLM5-crop on the global scale and demonstrated that it can successfully simulate the significant increases in global crop yield, which is consistent with our results. However, our study also provides a more comprehensive evaluation of CLM5-crop performance in China, which should prove useful for quantitatively assessing national and provincial food safety, given the good performance of CLM5-crop in simulating the national average and long-term trends of wheat, maize, and rice yield, and the temporal variability of yields in the top-10 producing provinces. Despite the above strengths of our study, there are two major limitations. First, we did not collect

yield statistics on scales smaller than the provincial scale (e.g., county scale), so we could not evaluate simulations of the spatial distribution of yield averages, trends, and interannual variability in more detail. And second, we did not evaluate the simulation of carbon fluxes. Evaluation of the carbon cycle may help in identifying the reasons for yield simulation skill.

## 5. Conclusions

The performances of 13 GGCMs (12 models from the GGCM phase 1 default scenario, and CLM5-crop) in simulating crop yields in China during 1980–2009 were comprehensively evaluated in this study. Results showed that the selected GGCM models tend to overestimate wheat, maize, and soybean yields but underestimate rice yield in China, with the MME result being 12%, 24%, and 41% higher than observed for wheat, maize, and soybean, and 25% lower for rice, while CLM5-crop can simulate the national yields of wheat, maize, and rice well. GGCMs generally fail to simulate the spatial pattern of crop yields in China. No model can simulate the national spatial distribution of wheat yield skillfully, and only one, one, and two models perform well for maize, rice, and soybean, respectively. GGCM models perform poorly in simulating the long-term trends of crop yields in China, but some can simulate the interannual variability and extremes reasonably well. Conversely, CLM5-crop can capture the observed significant upward trends for wheat, maize, and rice, but works poorly in simulating the interannual variability. In addition, at least one model can skillfully reproduce the observed temporal variability of yield in the top-10 producing provinces, except one province for wheat and two provinces for soybean. The best models are generally CLM4.5post-crop for wheat, CLM5-crop for maize and rice, and PEPIC or GGCM-MME for soybean.

There were a few limitations in our study. First, we focused on evaluating GGCMs at national and provincial scales, and the performances of GGCMs at smaller scales (e.g., county scale) remain unknown, which needs addressing with further collection and collation of yield statistics. In addition, the responses of yields to extreme weather and climate need to be evaluated in the near future. Lastly, the reasons behind simulation biases should be more thoroughly investigated by analyzing related processes (e.g., carbon cycle processes).

The findings of our study can serve as guidance for GGCM development in the future. Our results suggest that model development should target and consider cultivar differences and improvement and improve the representation of agronomic management (e.g., machinery and technological advances, multiple cropping seasons, residue management and soil tillage, and more realistic fertilizer and irrigation application). Also, parameter calibration and optimization as well as the representation of the influence of pests and extreme weather disasters on agricultural systems may improve the simulation skill for the interannual variability



of yields. In addition, our evaluation results could be helpful for understanding model results and biases, as well as model selection when predicting or projecting national and provincial yields, and for developing correction methods that can be applied to model yield outputs.

**Acknowledgements.** This study was co-supported by the Guangdong Major Project of Basic and Applied Basic Research (Grant No. 2021B0301030007), the National Key Research and Development Program of China (Grant Nos. 2017YFA0604302 and 2017YFA0604804), the National Natural Science Foundation of China (Grant No. 41875137), and the National Key Scientific and Technological Infrastructure project “Earth System Science Numerical Simulator Facility” (EarthLab). We thank Christoph Müller for his help in answering our question about GGCM phase 1 simulation data, the two anonymous reviewers for their valuable comments and suggestions, and the Editor for handling our paper.

**Electronic supplementary material:** Supplementary material is available in the online version of this article at <https://doi.org/10.1007/s00376-023-2234-3>.

## REFERENCES

- Balkovič, J., and Coauthors, 2014: Global wheat production potentials and management flexibility under the representative concentration pathways. *Global and Planetary Change*, **122**, 107–121, <https://doi.org/10.1016/j.gloplacha.2014.08.010>.
- Barlow, K. M., B. P. Christy, G. J. O’Leary, P. A. Riffkin, and J. G. Nuttall, 2015: Simulating the impact of extreme heat and frost events on wheat crop production: A review. *Field Crops Research*, **171**, 109–119, <https://doi.org/10.1016/j.fcr.2014.11.010>.
- Bondeau, A., and Coauthors, 2007: Modelling the role of agriculture for the 20th century global terrestrial carbon balance. *Global Change Biology*, **13**, 679–706, <https://doi.org/10.1111/j.1365-2486.2006.01305.x>.
- de Wit, A. J. W., and C. A. Van Diepen, 2008: Crop growth modelling and crop yield forecasting using satellite-derived meteorological inputs. *International Journal of Applied Earth Observation and Geoinformation*, **10**, 414–425, <https://doi.org/10.1016/j.jag.2007.10.004>.
- Deryng, D., W. J. Sacks, C. C. Barford, and N. Ramankutty, 2011: Simulating the effects of climate and agricultural management practices on global crop yield. *Global Biogeochemical Cycles*, **25**, GB2006, <https://doi.org/10.1029/2009GB003765>.
- Deryng, D., D. Conway, N. Ramankutty, J. Price, and R. Warren, 2014: Global crop yield response to extreme heat stress under multiple climate change futures. *Environmental Research Letters*, **9**, 034011, <https://doi.org/10.1088/1748-9326/9/3/034011>.
- Elliott, J., and Coauthors, 2014: The parallel system for integrating impact models and sectors (pSIMS). *Environmental Modelling & Software*, **62**, 509–516, <https://doi.org/10.1016/j.envsoft.2014.04.008>.
- Elliott, J., and Coauthors, 2015: The global gridded crop model intercomparison: Data and modeling protocols for phase 1 (v1.0). *Geoscientific Model Development*, **8**, 261–277, <https://doi.org/10.5194/gmd-8-261-2015>.
- FAO, 2021: *World Food and Agriculture—Statistical Yearbook 2020*. Food and Agriculture Organization of the United Nations, Rome, 353 pp, <https://doi.org/10.4060/cb1329en>. <https://doi.org/10.4060/cb1329en>.
- FAOSTAT, 2022: Food and agriculture data. [Available online from <https://www.fao.org/faostat/en/#home>]
- Folberth, C., T. Gaiser, K. C. Abbaspour, R. Schulin, and H. Yang, 2012: Regionalization of a large-scale crop growth model for sub-Saharan Africa: Model setup, evaluation, and estimation of maize yields. *Agriculture, Ecosystems & Environment*, **151**, 21–33, <https://doi.org/10.1016/j.agee.2012.01.026>.
- Franke, J. A., and Coauthors, 2020: The GGCM Phase 2 experiment: Global gridded crop model simulations under uniform changes in CO<sub>2</sub>, temperature, water, and nitrogen levels (protocol version 1.0). *Geoscientific Model Development*, **13**, 2315–2336, <https://doi.org/10.5194/gmd-13-2315-2020>.
- Ghose, B., 2014: Food security and food self-sufficiency in China: From past to 2050. *Food and Energy Security*, **3**, 86–95, <https://doi.org/10.1002/fes3.48>.
- Hantson, S., and Coauthors, 2020: Quantitative assessment of fire and vegetation properties in simulations with fire-enabled vegetation models from the Fire Model Intercomparison Project. *Geoscientific Model Development*, **13**, 3299–3318, <https://doi.org/10.5194/gmd-13-3299-2020>.
- Heinicke, S., K. Frieler, J. Jägermeyr, and M. Mengel, 2022: Global gridded crop models underestimate yield responses to droughts and heatwaves. *Environmental Research Letters*, **17**, 044026, <https://doi.org/10.1088/1748-9326/ac592e>.
- Hurt, G. C., and Coauthors, 2011: Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Climatic Change*, **109**, 117–161, <https://doi.org/10.1007/s10584-011-0153-2>.
- Jägermeyr, J., and Coauthors, 2020: A regional nuclear conflict would compromise global food security. *Proceedings of the National Academy of Sciences of the United States of America*, **117**, 7071–7081, <https://doi.org/10.1073/pnas.1919049117>.
- Jones, J. W., and Coauthors, 2003: The DSSAT cropping system model. *European Journal of Agronomy*, **18**, 235–265, [https://doi.org/10.1016/S1161-0301\(02\)00107-7](https://doi.org/10.1016/S1161-0301(02)00107-7).
- Keating, B. A., and Coauthors, 2003: An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, **18**, 267–288, [https://doi.org/10.1016/S1161-0301\(02\)00108-9](https://doi.org/10.1016/S1161-0301(02)00108-9).
- Kiniry, J. R., J. R. Williams, D. J. Major, R. C. Izaurralde, P. W. Gassman, M. Morrison, R. Bergentine, and R. P. Zentner, 1995: EPIC model parameters for cereal, oilseed, and forage crops in the northern Great Plains region. *Canadian Journal of Plant Science*, **75**, 679–688, <https://doi.org/10.4141/cjps95-114>.
- Kukul, M. S., and S. Irmak, 2018: Climate-driven crop yield and yield variability and climate change impacts on the U.S. Great Plains agricultural production. *Scientific Reports*, **8**, 3450, <https://doi.org/10.1038/s41598-018-21848-2>.
- Lawrence, D. M., and Coauthors, 2019: The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, **11**, 4245–4287, <https://doi.org/10.1029/2018MS001583>.
- Levis, S., A. Badger, B. Drewniak, C. Nevison, and X. L. Ren,

- 2018: CLMcrop yields and water requirements: Avoided impacts by choosing RCP 4.5 over 8.5. *Climatic Change*, **146**, 501–515, <https://doi.org/10.1007/s10584-016-1654-9>.
- Levis, S., G. B. Bonan, E. Kluzek, P. E. Thornton, A. Jones, W. J. Sacks, and C. J. Kucharik, 2012: Interactive crop management in the Community Earth System Model (CESM1): Seasonal influences on land–atmosphere fluxes. *J. Climate*, **25**, 4839–4859, <https://doi.org/10.1175/JCLI-D-11-00446.1>.
- Li, F., 2011: Probabilistic seasonal prediction of summer rainfall over East China based on multi-model ensemble schemes. *Acta Meteorologica Sinica*, **25**, 283–292, <https://doi.org/10.1007/s13351-011-0304-4>.
- Li, F., and Coauthors, 2019: Historical (1700–2012) global multi-model estimates of the fire emissions from the Fire Modeling Intercomparison Project (FireMIP). *Atmospheric Chemistry and Physics*, **19**, 12 545–12 567, <https://doi.org/10.5194/acp-19-12545-2019>.
- Li, Z. H., C. S. Zhan, S. Hu, L. K. Ning, L. F. Wu, and H. Guo, 2022: Evaluation of global gridded crop models (GGCMs) for the simulation of major grain crop yields in China. *Hydrology Research*, **53**, 353–369, <https://doi.org/10.2166/nh.2022.087>.
- Lindeskog, M., A. Arneth, A. Bondeau, K. Waha, J. Seaquist, S. Olin, and B. Smith, 2013: Implications of accounting for land use in simulations of ecosystem carbon cycling in Africa. *Earth System Dynamics*, **4**, 385–407, <https://doi.org/10.5194/esd-4-385-2013>.
- Liu, L. L., Y. Zhu, L. Tang, W. X. Cao, and E. L. Wang, 2013: Impacts of climate changes, soil nutrients, variety types and management practices on rice yield in East China: A case study in the Taihu region. *Field Crops Research*, **149**, 40–48, <https://doi.org/10.1016/j.fcr.2013.04.022>.
- Liu, W. F., H. Yang, C. Folberth, X. Y. Wang, Q. Y. Luo, and R. Schulin, 2016: Global investigation of impacts of PET methods on simulating crop–water relations for maize. *Agricultural and Forest Meteorology*, **221**, 164–175, <https://doi.org/10.1016/j.agrformet.2016.02.017>.
- Lobell, D. B., G. Bala, and P. B. Duffy, 2006: Biogeophysical impacts of cropland management changes on climate. *Geophys. Res. Lett.*, **33**, L06708, <https://doi.org/10.1029/2005GL025492>.
- Lombardozzi, D. L., Y. Q. Lu, P. J. Lawrence, D. M. Lawrence, S. Swenson, K. W. Oleson, W. R. Wieder, and E. A. Ainsworth, 2020: Simulating agriculture in the Community Land Model version 5. *J. Geophys. Res.: Biogeosci.*, **125**, e2019JG005529, <https://doi.org/10.1029/2019JG005529>.
- Luo, Y. C., Z. Zhang, Y. Chen, Z. Y. Li, and F. L. Tao, 2020: ChinaCropPhen1km: A high-resolution crop phenological dataset for three staple crops in China during 2000–2015 based on leaf area index (LAI) products. *Earth System Science Data*, **12**, 197–214, <https://doi.org/10.6084/m9.figshare.8313530>.
- Martre, P., and Coauthors, 2015: Multimodel ensembles of wheat growth: Many models are better than one. *Global Change Biology*, **21**, 911–925, <https://doi.org/10.1111/gcb.12768>.
- Müller, C., and Coauthors, 2017: Global gridded crop model evaluation: Benchmarking, skills, deficiencies and implications. *Geoscientific Model Development*, **10**, 1403–1422, <https://doi.org/10.5194/gmd-10-1403-2017>.
- Müller, C., and Coauthors, 2019: The global gridded crop model intercomparison phase 1 simulation dataset. *Scientific Data*, **6**, 50, <https://doi.org/10.1038/s41597-019-0023-8>.
- Portmann, F. T., S. Siebert, and P. Döll, 2010: MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochemical Cycles*, **24**, GB1011, <https://doi.org/10.1029/2008GB003435>.
- Ray, D. K., J. S. Gerber, G. K. MacDonald, and P. C. West, 2015: Climate variation explains a third of global crop yield variability. *Nature Communications*, **6**, 5989, <https://doi.org/10.1038/ncomms6989>.
- Rosenzweig, C., and Coauthors, 2014: Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 3268–3273, <https://doi.org/10.1073/pnas.1222463110>.
- Rötter, R. P., M. Appiah, E. Fichtler, K. C. Kersebaum, M. Trnka, and M. P. Hoffmann, 2018: Linking modelling and experimentation to better capture crop impacts of agroclimatic extremes—A review. *Field Crops Research*, **221**, 142–156, <https://doi.org/10.1016/j.fcr.2018.02.023>.
- Ruane, A. C., R. Goldberg, and J. Chrystanthacopoulos, 2015: Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation. *Agricultural and Forest Meteorology*, **200**, 233–248, <https://doi.org/10.1016/j.agrformet.2014.09.016>.
- Sinclair, T. R., and T. W. Ruffy, 2012: Nitrogen and water resources commonly limit crop yield increases, not necessarily plant genetics. *Global Food Security*, **1**, 94–98, <https://doi.org/10.1016/j.gfs.2012.07.001>.
- Sperber, K. R., H. Annamalai, I. S. Kang, A. Kitoh, A. Moise, A. Turner, B. Wang, and T. Zhou, 2013: The Asian summer monsoon: An intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century. *Climate Dyn.*, **41**, 2711–2744, <https://doi.org/10.1007/s00382-012-1607-6>.
- Sun, H. Y., X. Y. Zhang, S. Y. Chen, D. Pei, and C. M. Liu, 2007: Effects of harvest and sowing time on the performance of the rotation of winter wheat–summer maize in the North China Plain. *Industrial Crops and Products*, **25**, 239–247, <https://doi.org/10.1016/j.indcrop.2006.12.003>.
- Tian, S. Z., X. X. Dong, H. H. Guo, L. Dong, Y. F. Zhang, S. L. Liu, and J. F. Luo, 2019: Key soil nutrient requirements for different yield levels in North China. *Soil Science and Plant Nutrition*, **65**, 519–524, <https://doi.org/10.1080/00380768.2019.1639215>.
- Wallach, D., and Coauthors, 2018: Multimodel ensembles improve predictions of crop–environment–management interactions. *Global Change Biology*, **24**, 5072–5083, <https://doi.org/10.1111/gcb.14411>.
- Wang, E. L., Q. Yu, D. R. Wu, and J. Xia, 2008: Climate, agricultural production and hydrological balance in the North China Plain. *International Journal of Climatology*, **28**, 1959–1970, <https://doi.org/10.1002/joc.1677>.
- Wu, J., and X.-J. Gao, 2013: A gridded daily observation dataset over China region and comparison with the other datasets. *Chinese Journal of Geophysics*, **56**, 1102–1111, <https://doi.org/10.6038/cjg20130406>. (in Chinese with English abstract)
- Wu, X., and Coauthors, 2016: ORCHIDEE-CROP (v0), a new process-based agro-land surface model: Model description and evaluation over Europe. *Geoscientific Model Development*, **9**, 857–873, <https://doi.org/10.5194/gmd-9-857-2016>.
- Xiao, D. P., and F. L. Tao, 2014: Contributions of cultivars, management and climate change to winter wheat yield in the

- North China Plain in the past three decades. *European Journal of Agronomy*, **52**, 112–122, <https://doi.org/10.1016/j.eja.2013.09.020>.
- Yin, Y., Q. Tang, and X. Liu, 2015: A multi-model analysis of change in potential yield of major crops in China under climate change. *Earth System Dynamics*, **6**, 45–59, <https://doi.org/10.5194/esd-6-45-2015>.
- Yu, Y. Q., Y. Huang, and W. Zhang, 2012: Changes in rice yields in China since 1980 associated with cultivar improvement, climate and crop management. *Field Crops Research*, **136**, 65–75, <https://doi.org/10.1016/j.fcr.2012.07.021>.
- Zhao, H., and Coauthors, 2021: China's future food demand and its implications for trade and environment. *Nature Sustainability*, **4**, 1042–1051, <https://doi.org/10.1038/s41893-021-00784-6>.