The statistical emulators of GGCMI phase 2: responses of year-to-year variation of crop yield to CO_2, temperature, water and nitrogen perturbations

Weihang Liu[1,2,3,4], Tao Ye[1,2,3,4]*, Christoph Müller[5], Jonas Jägermeyr[5,6,7], James A. Franke[8,9], Haynes Stephens[8,9], Shuo Chen[1,2,3,4]

[1] State Key Laboratory of Earth Surface Processes and Resource Ecology (ESPRE), Beijing Normal University, Beijing, 100875, China

[2] Key Laboratory of Environmental Change and Natural Disasters, Ministry of Education, Beijing Normal University, Beijing 100875, China

[3] Academy of Disaster Reduction and Emergency Management, Ministry of Emergency Management and Ministry of Education, Beijing 100875, China

[4] Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

[5] Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany

[6] NASA Goddard Institute for Space Studies, New York City, New York, USA

[7] Center for Climate Systems Research, Columbia University, New York City, New York, USA

[8] Department of the Geophysical Sciences, University of Chicago, Chicago, Illinois, USA

[9] Center for Robust Decision- making on Climate and Energy Policy (RDCEP), University of Chicago, Chicago, Illinois, USA

*** Corresponding author at:**

State Key Laboratory of Earth Surface Processes and Resource Ecology (ESPRE), Beijing Normal University, 19 Xinjiekouwai Street, Beijing, 100875, China.

E-mail: yetao@bnu.edu.cn

Abstract

Understanding the impact of climate change on year-to-year variation of crop yield is critical to global food stability and security. While crop model emulators are believed to be lightweight tools to replaces the raw models~~ per se~~, few emulators have been developed to capture such interannual variation of crop yield in response to climate variability. In this study, we developed a statistical emulator with machine learning algorithm to reproduce the response of year-to-year variation of four crop yield to CO_2 (C), temperature (T), water (W) and nitrogen (N) perturbations defined in the Global Gridded Crop Model Intercomparison Project (GGCMI) phase 2 experiment. The emulators were able to explain more than ~~92~~52% variance of simulated yield and performed well in capturing the year-to-year variation of global average and gridded crop yield over current croplands in the baseline. With the changes in CTWN perturbations, the emulators could well reproduce the year-to-year variation of crop yield over most current cropland. The variation of *R* and the mean absolute error was small under the single CTWN perturbations and dual factor perturbations. These emulators thus provide statistical response surfaces of yield, including both its mean and interannual variability, to climate factors. They could facilitate spatiotemporal downscaling of crop model simulation, projecting the changes in crop yield variability in the future, and serving as a lightweight tool of multi-model ensemble simulation. The emulators enhanced the flexibility of crop yield estimates and expanded the application of large-ensemble simulation of crop yield under climate change.

1. Introduction

The impact of climate change on crop yield is an increasing concern of global food security (Kinnunen et al., 2020). Two distinct approaches have been used to evaluate the impact of climate change on crop yield, process-based crop models and statistical models. Process-based crop models are reliable tools to project crop yields under future climate change but computationally expensive (Jones et al., 2017). In contrast, statistical models are lightweight tools that could fit yield response to historical climate change (Li et al., 2019b) but the relationship between climate factors and crop yield is ~~limited by~~based on the ~~current~~ historical climate conditions and their effects on crop yields, which can hardly be used for future projection with new, unprecedented climate conditions. Therefore, it is promising to develop tools that can reduce the expense of computation and increase capacity for flexible future projections (Franke et al., 2020a).

Earlier studies have developed statistical emulators of process-based crop model results to balance the advantages and disadvantages of process-based crop models and statistical models. Those statistical emulators were initially developed with "entire scenarios" (simultaneous changes in climate factors) simulation during historical or future periods. Emulators have been developed for process-based crop models, like APSIM (Shahhosseini et al., 2019), GEPIC (Folberth et al., 2019), GWG (Xu et al., 2021), GAZE (Raimondo et al., 2021), and WOFOST (Tartarini et al., 2021), and used to estimate historical crop yield. As the emulator trained by historical simulation could not project the crop yield in the future, multiple crop model ensemble simulation in future climate scenarios were used to calibrate emulators (Blanc, 2017, 2020; Blanc and Sultan, 2015; Mistry et al., 2017; Ostberg et al., 2018). However, t~~T~~he scenario-based future crop yield projection is not a systematic perturbation of climate factors ~~change~~ (Franke et al., 2020a). For instance, the scenario-based yield projection can only provide the simulated crop yield driven by simultaneous changes in climate factors. The dependency of temperature and precipitation will be kept in scenarios, such that the impact of temperature and precipitation cannot be clearly separated.

An alternative emulation based on "perturbated factors" training dataset was introduced, which offers advantages to separate effects of crop yield drivers. The perturbated factors emulation was first conducted on site-based crop model simulations, which could estimate the yield across a broad range of CO_2, temperature and water (Fronzek et al., 2018; Makowski et al., 2015; Pirttioja et al., 2015) but these emulators were limited to the site-level. To break the constrain of site-based simulation, the global gridded crop model intercomparison (GGCMI) phase 2 provided a simulation dataset across structured CO_2-Temperature-Water-Nitrogen (CTWN) perturbation cubes. This dataset offered two major advantages: it allows for separating the effects of different climatic factors and of nitrogen levels on crop yields, and to distinguish the climatological-mean and year-to-year variation of yields (Franke et al., 2020b). The phase 2 dataset was published to support the derivation of crop yield- climate change "response surfaces". Based on the CTWN cubes, a statistical emulator has been developed providing near-global-coverage multi-model emulators of climatological-mean yield projections from the GGCMI Phase 2 ensemble by using a regression model with a third-order polynomial basis function (Franke et al., 2020a). Due to the focus on climatological-mean yield, the aspect of year-to-year variation of yield under CTWN perturbations has not been fully analyzed or exploited in emulator design.

For climate change risk assessment, interannual yield variability (or the year-to-year variation of yield) is an important metric of yield risk (Liu et al., 2021b) and food supply stability (Liu et al., 2021a) but has been insufficiently addressed in previous studies (Campbell et al., 2016). Large year-to-year variation of crop yield can influence livelihoods of producers, food prices (Hasegawa et al., 2021), hunger (Janssens et al., 2020) and even lead to political instabilities (Sternberg, 2011). Recently, year-to-year variation has been introduced as a metric for climate change risk on global crop production (Jägermeyr et al., 2021). Developing statistical emulators that can reproduce the year-to-year variation of yield from the CTWN cubes could therefore provide a powerful tool for studies focusing on the risk of climate change impact on yield. In this study, we aimed exclusively to develop statistical emulators to reproduce year-to-year yield variation with GGCMI phase 2 experiment data.

2. Data and Methods

2.1 Data

The input and output data for the simulation of global gridded crop yield were obtained from the GGCMI phase 2 experiment dataset, which includes gridded crop yield projections at 0.5° longitudinal/latitudinal resolution for maize, spring wheat, winter wheat, and rice~~, and soybean~~ (Franke et al., 2020b). The input data for the process-based simulations in GGCMI Phase 2 included data of climate, soil, atmospheric CO_2 concentration, and nitrogen fertilizer application rates. Baseline (1980-2010) climate inputs were used from the AgMIP Modern-Era Retrospective Analysis for Research and Applications (AgMERRA) forcing dataset, including daily maximum and minimum temperatures, precipitation, and solar radiation (Ruane et al., 2015). Systematic perturbations were conducted in each grid cell with seven temperature levels (from -1 K to +6 K in 1K interval, with +5K skipped), nine precipitation levels (from -50% to +30%, in 10% interval, with -40% skipped, the Winf precipitation level is simulation under fully irrigated condition), four CO_2-concentration levels (360, 510, 660, and 810 ppm), and three nitrogen levels (10, 60, and 200 kg/ha). Simulations were repeated for two adaptation strategies, i.e. no adaptation in cultivar (A0) and adaptation by maintaining growing season length (A1). Twelve GGCMs were then forced with each of these perturbations of the original reanalysis weather data. We selected 10 of 12 crop models in the GGCMI phase 2 experiment for constructing the emulators. These were APSIM-UGOE, CARAIB, EPIC-IIASA, EPIC-TAMU, GEPIC, LPJ-GUESS, LPJmL, ORCHIDEE-crop, pDSSAT, and PEPIC (**Table1**). PROMET and JULES were not included as they used different climate inputs.

The GGCMs used a national and subnational crop calendar for crops that is based on Sacks et al (2010), Portmann et al (2010), and environment-based extrapolations (Elliott et al., 2015). The crop calendar was used to determine the window to calculate the climatic predictors and grid-specific growing season length. The current global harvested area for identifying currently used cropland was obtained from the spatial production allocation model (SPAM) whose spatial resolution was 10km. The soil type data was obtained from the Harmonized World Soil Database (Nachtergaele et al., 2009).

Table 1 GGCMs included in emulation. Each model offers the same set of CTWN simulations across four crops.

GGCMs	Maize	Winter wheat	Spring wheat	Rice
APSIM-UGOE	√	√	√	√
CARAIB	√	√	√	√
EPIC-IIASA	√	√	√	√
EPIC-TAMU	√	√	√	√
GEPIC	√	√	√	√
LPJ-GUESS	×	√	√	×
LPJmL	√	√	√	√
ORCHIDEE-crop	√	√	×	√
pDSSAT	√	√	√	√
PEPIC	√	√	√	√

* LPJ-GUESS omits maize and rice, and ORCHIDEE-crop omits spring wheat (denoted by "×")

2.2 Methods

Our study focused on the development and evaluation of emulators, which contains the following steps: 1) defining the predictors used to train the emulators; 2) preparing the predictors with climatic and non-climatic data; 3) training and cross validating the emulators with machine learning algorithm; and 4) evaluating the performance of emulators (**Figure1**).

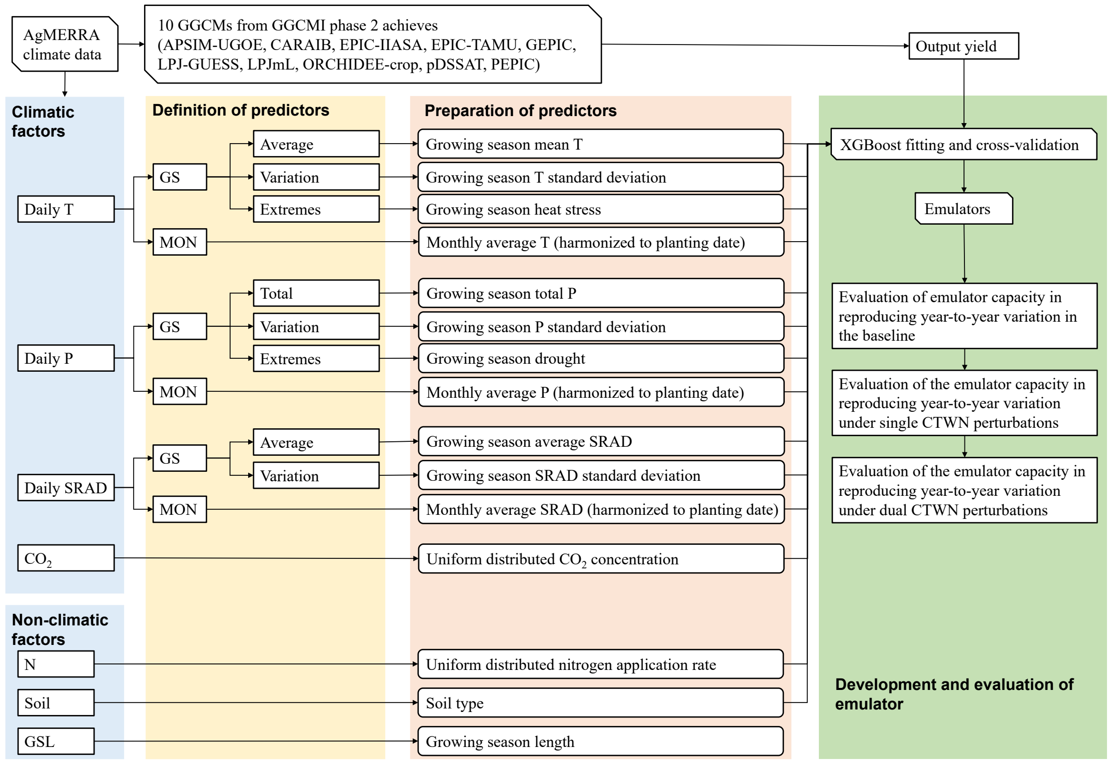

Figure 1 Overall framework of emulator development for GGCMs. Each GGCM-crop combination was calibrated as an emulator independently. T: temperature, processed separately for daily maximum, and minimum temperatures, P: precipitation, SRAD: solar radiation, N: nitrogen, Soil: soil properties. When developing irrigated yield emulator, the precipitation-related predictors are excluded.

2.2.1 Definition and preparation of predictors

All the predictors were computed or adapted from the GGCMs' input and output datasets. The climatic predictors were defined at two time-scales, growing season (GS) and monthly (MON) (**Table2**). The growing season average temperature, total precipitation and average solar radiation were able to explain the variation of climatological mean yield of GGCM phase2 (Franke et al., 2020a). To improve the capacity of emulators in reproducing the year-to-year variation of crop model yield, daily variability and extremes of climate factors during the growing seasons were considered here. The variation of temperature, precipitation and solar radiation during the growing seasons were calculated with the standard deviation of their daily values in each growing season, which represents the intensity of daily fluctuation of weather. Additionally, the heat and drought were selected to be the extreme climate predictors, which was quantified by extreme degree day (EDD, cumulative temperature that exceed the high temperature threshold, Lobell et al., 2012) and maximum consecutive ~~drought~~ dry days (CDD, maximum length of consecutive days without precipitation, Troy et al., 2015), because the negative effect of these two extremes could be shown by the current GGCM (Heinicke et al., 2022). Other climate extremes, like excessive wetness, was not used because the GGCM failed to show the negative effect (Li et al., 2019a; Liu et al., 2022).

The monthly predictors only consisted of monthly average values. The monthly average temperature, total precipitation and average solar radiation were harmonized according to the specific planting date. The number of months was determined with the crop-specific maximum growing season length over the global cropland defined by GGCMI phase2 experiment. For winter and spring wheat, we prepared the climatic predictors over 10 and eight months after sowing. For maize and rice, climatic predictors over eight and seven months after sowing were used, respectively.

The atmospheric CO_2 concentration and the nitrogen application rate were uniformly distributed predictors. All years and grid cells were set at the same CO_2 concentration and nitrogen application rate for each perturbation. Soil property is an important temporally constant predictor, whose interaction with climate played important role in yield simulation and emulator development (Blanc, 2017). As the soil parameter settings of each GGCM varied, we selected the soil type at each grid to represent the spatial variation of soil properties. There were 13 soil types, including heavy clay, silty clay, light clay, silt clay loam, clay loam, silt, silt loam, sandy clay, loam, sandy clay loam, sandy loam, loamy sand, sand. The most obvious difference across cultivars over the global croplands is the growing degree requirement to reach maturity, which was determined by both mean climatology and cultivar traits. To reproduce the length of days from planting date to maturity date given by GGCMI phase2 crop calendar input~~spatial difference of simulated crop yield~~, we added a temporal constant growing season length~~spatial difference term~~ as a predictor, i.e. temporal constant growing season length.

As the purpose of emulator training is to develop a lightweight tool for crop simulation, there has always been a trade-off between the goodness-of-fit and the number of predictors. Therefore, we considered three strategies of using our predictors. “Strategy A” uses all predictors (the “Full” model), which is expected to derive the best goodness-of-fit. “Strategy B” uses only climatic predictors during growing season scale (the “GS” model), together with CO_2 concentration, nitrogen application rate and site information, soil class and growing season length. “Strategy C” uses only monthly average climatic predictors with other location-invariant predictors (the “Mon” model). In general, strategy B uses the smallest number of predictors, but those predictors need to be computed from daily climate forcing. Stagey C only relays on monthly climate data, and therefore is the least costly strategy for data preparation. A comparison between the three strategies would help us find a good balance between the predictors used and overall goodness-of-fit of the emulators.

Table 2 Predictors of emulation. For rainfed yield emulators, we used all these predictors but for fully-irrigated yield emulators, the precipitation predictors were not included. Full, GS and Mon were three strategies to develop emulators. Full: developing emulators with all the climatic predictors; GS: developing emulators with climatic predictors during growing season scale; Mon: developing emulators with climatic predictors during monthly scale.

Predictor abbreviations	Descriptions	References	Full	GS	Mon	Time
	Temperature related predictors					
$GDD_{low\text{-}high}$_GS	Growing degree day during growing season (winter wheat: low=0°C, high=30°C; spring wheat: low=5°C, high=30°C; maize: low=8°C, high=30°C; rice: low=10°C, high=35°C)	(Frieler et al., 2017; Jägermeyr et al., 2020; Lobell et al., 2012)				1
EDD_{high+}_GS	Extreme degree day during growing (winter and spring wheat, maize: high=30°C; rice: high=35°C	(Lobell et al., 2012)				1
Tmax_GSmean	Average daily maximum temperature during growing season	(Zhu and Troy, 2018)				1

Tmin_GSmean	Average daily minimum temperature during growing season	(Zhu and Troy, 2018)	1
Tmax_GSstd	Standard deviation of daily maximum temperature during growing season	(Zhu and Troy, 2018)	1
Tmin_GSstd	Standard deviation of daily minimum temperature during growing season	(Zhu and Troy, 2018)	1
Tmax_MONmean	Harmonized monthly average daily maximum temperature (MON=1–10 for winter wheat, MON=1–8 for spring wheat and maize, MON=1–7 for rice, since planting date)	(Folberth et al., 2019) (Jägermeyr et al., 2020)	1
Tmin_MONmean	Harmonized monthly average daily minimum temperature (MON=1–10 for winter wheat, MON=1–8 for spring wheat and maize, MON=1–7 for rice, since planting date)	(Folberth et al., 2019) (Jägermeyr et al., 2020)	1
Precipitation related predictors			
Pre_GSsum	Total daily precipitation during growing season	(Troy et al., 2015)	1
Pre_GSstd	Standard deviation of daily precipitation during growing season	(Zhu and Troy, 2018)	1
CDD_GS	Consecutive drought day (daily precipitation=0)	(Troy et al., 2015)	1
Pre_MONsum	Harmonized monthly total precipitation (MON=1–10 for winter wheat, MON=1–8 for spring wheat and maize, MON=1–7 for rice, since planting date)	(Folberth et al., 2019) (Jägermeyr et al., 2020)	1
Solar radiation related predictors			
SRAD_GSmean	Average daily solar radiation during growing season	(Folberth et al., 2019)	1
SRAD_GSstd	Standard daily solar radiation during growing season	(Folberth et al., 2019)	1
SRAD_MONmean	Harmonized monthly average daily solar radiation (MON=1–10 for winter wheat, MON=1–8 for spring wheat and maize, MON=1–7 for rice, since planting date)	(Folberth et al., 2019) (Jägermeyr et al., 2020)	1
Greenhouse gas concentration			
CO_2	CO_2 concentration	(Franke et al., 2020a)	2
Non-climatic predictors			
N	Nitrogen fertilizer application	(Franke et al., 2020a)	2

Soil_type	Soil type	(Blanc, 2017)				3
~~SDT~~GSL	~~Spatial difference term~~Growing season length	(Folberth et al., 2019)				3

*The colored the row denotes the predictors was included in the emulator. The column "Time" is defined to clarify the spatiotemporal dynamics of predictors: "1" represents both time and space variant predictors, "2" represents space invariant predictors, "3" represents time invariant predictors.

2.2.2 Emulator training and validation

Training the emulator of specific GGCM is to derive the response relationship between input and output, so that the emulator could replicate the complex process of yield simulation within the crop model. Emulation was trained by using machine learning regression on the GGCMI-2 ensemble of crop- specific simulated yield with all CTWN perturbations. Each grid-year-perturbation combination was regarded as a sample in the fitting. We developed emulators of irrigated and rainfed yield and in A0 and A1 scenarios separately. Since the outputs of GGCM outside the current croplands were not well examined, we trained the machine learning based emulators only on currently used cropland, according to the SPAM data for each crop separately.

The extreme gradient boosting (XGBoost) algorithm, a highly efficient realization of the gradient boosting approach that showed the best performance in recent machine learning challenges (Chen and Guestrin, 2016), was used to train the emulators. Key parameters in XGBoost, including the learning rate (0.1), the number of estimators (4000), and the maximum tree depths (10), were tuned by a grid search along parameter dimensions based on the default parameter as reference (Folberth et al., 2019). The goodness-of-fit of XGBoost was validated with the coefficient of determination R^2_{adjust}.

$$R^2_{adjust} = 1 - \frac{(n-1)\times(1-R^2)}{n-k}$$

where n is the sample size of the validation set, k is the number of predictors.

Considering the spatiotemporal autocorrelation of simulated crop yield given by GGCM, we now used a "held out years and regions" strategy for leave one-year-out cross-validation (Roberts et al., 2017; Sweet et al., 2023). Specifically, the all grid-year samples are split into N folds. N is determined by the number of Köppen–Geiger (KG) classes, which have more than 100 grid cells with harvested areas. If there are too few harvested areas in one KG class, it will not be included in the cross-validation process. For each fold of emulator training and validation, we withhold 10% of years (the last 3 years) and one entire KG class for validation, and the other grid-year samples are used for training the emulator. We think selecting continuous years for validation can avoid temporal autocorrelation. If we randomly select 10% of years, the correlation between adjacent years still exist. Actually, any continuous three years are able to solve this problem, such that we just use the last years according to the choice of (Sweet et al., 2023).~~We used two validation strategies to show the goodness-of-fit. Firstly, we used a 10-fold cross-validation that the samples were randomly divided into 10 folds, with nine of them used for training, and the rest used for validation. Secondly, considering the spatial autocorrelation in the raw GGCM simulated yield can invalidate the machine learning random selected validation sets (Ploton et al., 2020), we used the Köppen–Geiger climate regions to split the trained and validated sets. We used a leave-one-out approach that 29 out of the 30 climate regions were used for training, and the rest used for validation. The climate regions which contain less than 50 grids under current harvested areas will be removed from leave-one-out cross validation process.~~ Emulators were trained in Python3.8 with GPU (https://xgboost.readthedocs.io/en/latest/python/index.html).

2.2.3 Evaluation of emulator

Emulator performance was evaluated by comparing the 30-year emulated yield with the 30-year simulated yield of the GGCM. As we aimed at developing emulator that could replicate the year-to-year variation of yield, the correlation coefficient (*R*), mean absolute error (*MAE*) and mean relative error (*MRE*) were used to evaluate the performance of emulators in the baseline and varied perturbations.

$$R = \frac{\sum_{i=1}^{n}(Y_{XGB,i} - \bar{Y}_{XGB})(Y_{GGCM,i} - \bar{Y}_{GGCM})}{\sqrt{\sum_{i=1}^{n}(Y_{XGB,i} - \bar{Y}_{XGB})^2 \cdot (Y_{GGCM,i} - \bar{Y}_{GGCM})^2}}$$

$$MAE = \frac{\sum_{i=1}^{n}| Y_{XGB,i} - Y_{GGCM,i} |}{n}$$

$$MRE = \frac{\sum_{i=1}^{n}| (Y_{XGB,i} - Y_{GGCM,i}) / Y_{GGCM,i} |}{n}$$

where n is the sample size of the validation set, $Y_{GGCM,i}$ is the annual simulated yield of the GGCMs, $Y_{XGB,i}$ is the annual projected yield of the XGB algorithm, and $\bar{Y}_{XGB}$ and $\bar{Y}_{GGCM}$ were the average XGBoost predicted and GGCM simulated yield, respectively.

3. Results

3.1 Goodness-of-fit of the emulators training

Overall, the emulator developed with XGBoost algorithm could well reproduce the variance of GGCM yield simulations, with adjusted R^2 greater than ~~0.92~~0.52 (**Table3**). ~~The scatter plots of emulated yield and GGCM simulated yield for testing samples are clustered closely around the 1:1 ratio line (Figure S1).~~ For most emulators the adjusted R^2 under fully-irrigated (Winf) simulation were greater than those under rainfed simulation (W). Under A0 and A1 scenarios (The A0 denotes no adaptation and A1 denotes adaptation of the growing season to regain the original growing season length under warming scenarios that otherwise lead to accelerated phenology and thus shorter growing seasons.), the adjusted R^2 was comparable. For different crops, the performance of emulators developed for winter and spring wheat were slightly better than those developed for maize and rice. Among the GGCMs, ~~EPIC-IIASA~~PEPIC's behavior can best be emulated by emulators, with greatest R^2 values for all crops and scenarios. There are also several GGCM that is bit challenging for the XGB algorithm to capture, i.e. winter wheat and ~~spring wheat~~rice simulation from ~~GEPIC~~ORCHIDEE-crop, maize of pDSSAT, and ~~rice~~spring wheat of ~~EPIC-TAMU~~LPJmL, with R^2 values ranging from ~~0.92~~0.52 to ~~0.96~~0.63. ~~When using the Köppen–Geiger climate regions to apply the leave-one-out cross validation, the adjusted R^2 is generally smaller than those obtained by the 10-fold cross validation with randomly selected samples, ranging between 0.93 and 0.64 (Table S1).~~

~~**Table 3** Adjusted R^2 of XGBoost derived from 10-fold cross validation with randomly selected samples~~

~~GGCMs (A0)~~	~~Winter wheat~~		~~Spring wheat~~		~~Maize~~		~~Rice~~	
	~~Winf~~	~~W~~	~~Winf~~	~~W~~	~~Winf~~	~~W~~	~~Winf~~	~~W~~
~~APSIM-UGOE~~	~~0.99~~	~~0.97~~	~~0.98~~	~~0.97~~	~~0.92~~	~~0.94~~	~~0.96~~	~~0.96~~
~~CARAIB~~	~~0.98~~	~~0.98~~	~~0.98~~	~~0.98~~	~~0.97~~	~~0.97~~	~~0.98~~	~~0.97~~
~~EPIC-IIASA~~	~~0.99~~	~~0.98~~	~~0.99~~	~~0.99~~	~~0.99~~	~~0.98~~	~~0.99~~	~~0.98~~
~~EPIC-TAMU~~	~~0.97~~	~~0.97~~	~~0.97~~	~~0.97~~	~~0.94~~	~~0.97~~	~~0.93~~	~~0.97~~
~~GEPIC~~	~~0.97~~	~~0.95~~	~~0.98~~	~~0.96~~	~~0.97~~	~~0.95~~	~~0.97~~	~~0.96~~
~~LPJ-GUESS~~	~~0.99~~	~~0.98~~	~~0.99~~	~~0.98~~	-	-	-	-
~~LPJmL~~	~~0.98~~	~~0.98~~	~~0.98~~	~~0.98~~	~~0.94~~	~~0.96~~	~~0.95~~	~~0.96~~
~~ORCHIDEE-crop~~	~~0.99~~	~~0.98~~	-	-	~~0.98~~	~~0.96~~	~~0.97~~	~~0.97~~
~~pDSSAT~~	~~0.97~~	~~0.97~~	~~0.99~~	~~0.98~~	~~0.92~~	~~0.92~~	~~0.95~~	~~0.96~~
~~PEPIC~~	~~0.98~~	~~0.97~~	~~0.98~~	~~0.98~~	~~0.98~~	~~0.97~~	~~0.97~~	~~0.97~~
~~GGCMs (A1)~~	~~Winter wheat~~		~~Spring wheat~~		~~Maize~~		~~Rice~~	
	~~Winf~~	~~W~~	~~Winf~~	~~W~~	~~Winf~~	~~W~~	~~Winf~~	~~W~~
~~APSIM-UGOE~~	~~0.99~~	~~0.96~~	~~0.98~~	~~0.96~~	~~0.96~~	~~0.92~~	~~0.97~~	~~0.96~~
~~CARAIB~~	~~0.98~~	~~0.98~~	~~0.99~~	~~0.99~~	~~0.97~~	~~0.97~~	~~0.98~~	~~0.97~~
~~EPIC-IIASA~~	-	-	-	-	-	-	-	-
~~EPIC-TAMU~~	~~0.98~~	~~0.98~~	~~0.97~~	~~0.97~~	~~0.95~~	~~0.97~~	~~0.94~~	~~0.97~~
~~GEPIC~~	~~0.98~~	~~0.96~~	~~0.98~~	~~0.97~~	~~0.98~~	~~0.97~~	~~0.98~~	~~0.97~~
~~LPJ-GUESS~~	~~0.99~~	~~0.99~~	~~0.99~~	~~0.99~~	-	-	-	-
~~LPJmL~~	~~0.98~~	~~0.99~~	~~0.97~~	~~0.98~~	~~0.94~~	~~0.96~~	~~0.97~~	~~0.96~~
~~ORCHIDEE-crop~~	-	-	-	-	-	-	-	-
~~pDSSAT~~	~~0.97~~	~~0.97~~	~~0.98~~	~~0.98~~	~~0.93~~	~~0.92~~	~~0.96~~	~~0.96~~
~~PEPIC~~	~~0.98~~	~~0.97~~	~~0.99~~	~~0.98~~	~~0.98~~	~~0.97~~	~~0.97~~	~~0.97~~

Table 3 Adjusted R^2 of XGBoost derived from 10-fold cross validation with randomly selected samples

GGCMs (A0)	Winter wheat		Spring wheat		Maize		Rice	
	Winf	W	Winf	W	Winf	W	Winf	W
APSIM-UGOE	0.87	0.75	0.67	0.62	0.60	0.58	0.65	0.56
CARAIB	0.63	0.63	0.73	0.73	0.69	0.58	0.61	0.60

EPIC-IIASA	0.68	0.61	0.70	0.68	0.67	0.69	0.71	0.63
EPIC-TAMU	0.65	0.70	0.80	0.61	0.77	0.68	0.67	0.59
GEPIC	0.83	0.62	0.77	0.67	0.84	0.74	0.79	0.67
LPJ-GUESS	0.84	0.84	0.81	0.68	-	-	-	-
LPJmL	0.63	0.69	0.59	0.68	0.65	0.73	0.65	0.64
ORCHIDEE-crop	0.59	0.56	-	-	0.62	0.78	0.52	0.71
pDSSAT	0.63	0.60	0.69	0.65	0.55	0.51	0.63	0.58
PEPIC	0.80	0.78	0.90	0.75	0.85	0.75	0.79	0.71

GGCMs (A1)	Winter wheat		Spring wheat		Maize		Rice	
	Winf	W	Winf	W	Winf	W	Winf	W
APSIM-UGOE	0.85	0.73	0.69	0.64	0.60	0.59	0.62	0.56
CARAIB	0.59	0.58	0.73	0.71	0.64	0.53	0.71	0.68
EPIC-IIASA	-	-	-	-	-	-	-	-
EPIC-TAMU	0.67	0.61	0.76	0.64	0.81	0.63	0.68	0.60
GEPIC	0.91	0.69	0.83	0.71	0.88	0.79	0.90	0.87
LPJ-GUESS	0.94	0.87	0.87	0.72	-	-	-	-
LPJmL	0.69	0.71	0.57	0.68	0.71	0.79	0.61	0.60
ORCHIDEE-crop	-	-	-	-	-	-	-	-
pDSSAT	0.67	0.64	0.75	0.69	0.63	0.58	0.69	0.63
PEPIC	0.80	0.76	0.90	0.75	0.88	0.77	0.86	0.77

"-": No GGCM simulation; Winf: irrigated condition; W: rainfed condition. The A0 denotes no adaptation and A1 denotes cultivar adaptation to regain original growing season length under warming scenarios.

The adjusted R^2 of emulators developed with all predictors ("Full model") was greater than those developed with growing season predictors ("GS model") and monthly predictors ("MON model") (**Figure 2**). ~~If we validated using the 10-fold randomly selected sample approach, the performances of "GS model" and "MON model" were good and comparable to the "Full Model". Nevertheless, the difference in performance became much pronounced when validated by using the climate-zone based leave-one-out approach (Figure S2).~~ GS models would suffer from reduced number of predictors and their adjusted R^2s were 0.1~0.15 smaller than corresponding MON models. Still, Full models had the largest adjusted R^2 at the cost of the largest number of predictors. For later usage of the emulators, a trade-off must be taken between cost of preparing predictors and model goodness-of-fit, and the "MON model" could be a balanced choice as it required only monthly average weather conditions.

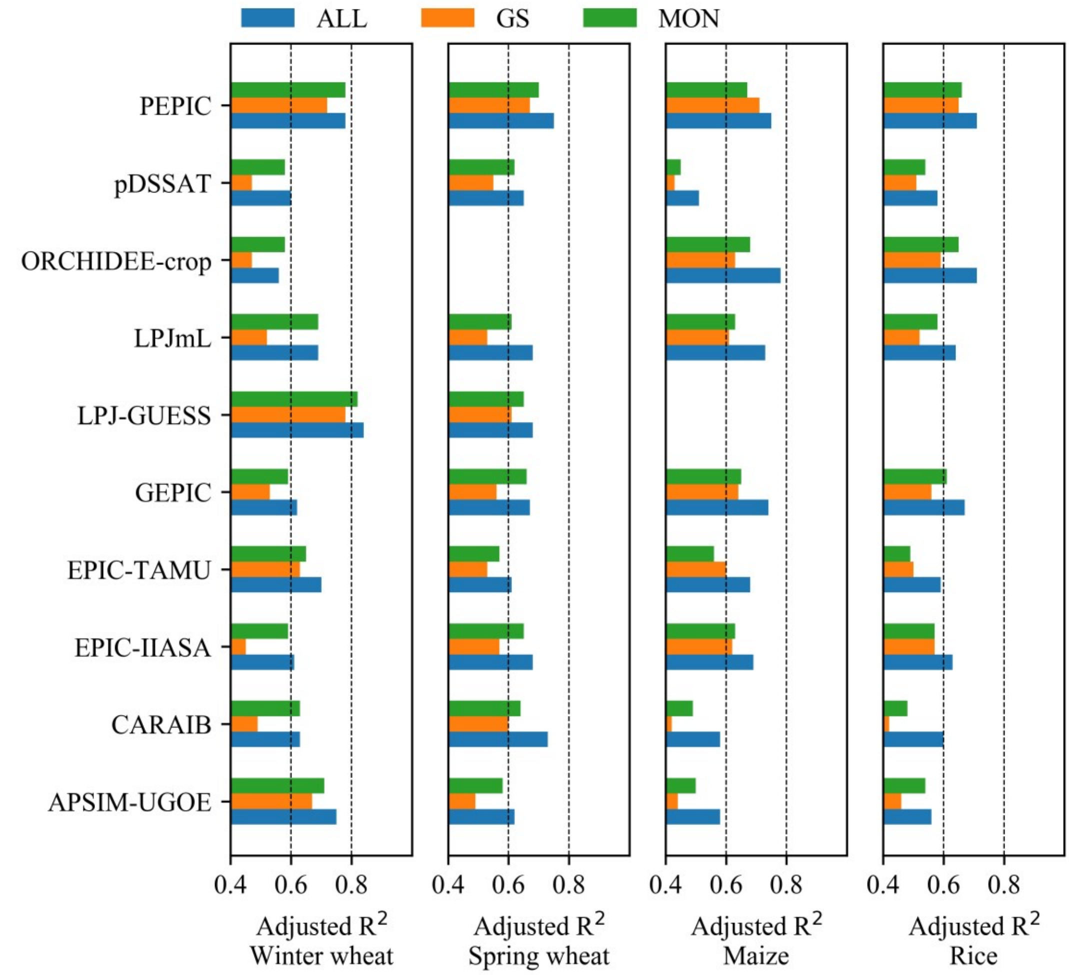

Figure 2 Adjusted R^2 of emulators (10-fold cross validation with randomly selected samples) with different strategery of predictors. All: “Full model”, GS: “GS model”, MON~~on~~: “~~MON~~ Mon model”. Emulators for ORCHIDEE by spring wheat, and LPJ-GUESS by Maize and Rice were not fitted due to the lack of simulation of raw GGCM.

3.2 Performance of emulators to capture the year-to-year variation of GGCM yield in the baseline

3.2.1 Performance of individual emulators at the global scale

Over current global cropland, the emulator of each GGCM could well reproduce the year-to-year variation of global average yield in the baseline period (during 1981–2010) (**Figure 3**). All individual emulators could capture the corresponding GGCM simulated yield, with scatters concentrated in the 1:1 ratio line. Different GGCM simulated yield levels varied from 1.7 to 7.8 t/ha but the performance of emulators has not been influenced.

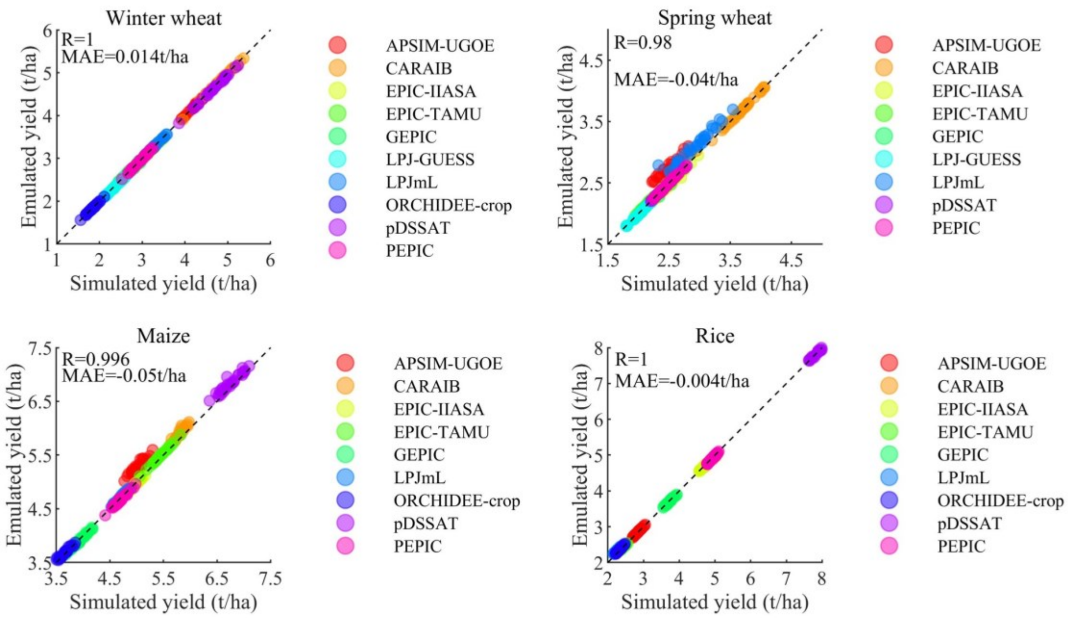

Figure 3 Emulator performance to reproduce the year-to-year variation of global average yield (1981 – 2010) over current cropland. As ORCHIDEE-crop has not simulated yield under C360T0W0N200, we used the C360T0W10N200 as the baseline. Each point with the same color is yield in 30 year. R is correlation coefficient and MAE is mean absolute error.

3.2.2 Performance of individual emulators at grid scale

The overall performances of emulators at grid level were good for most crop-GGCM combinations in the baseline. The performance of each emulator over current global cropland grids were plotted by using scatter of MAE and *R* (**Figure 4**). The capacity of emulators in reproducing the wheat yield simulated by GGCMs was better than that of maize and rice. The median *R* over current winter and spring wheat harvested areas were greater than ~~0.9~~0.7. The *R* of the EPIC-TAMU-emulator and the LPJ-GUESS-emulator were relatively smaller than other eight emulators developed for winter and spring wheat, respectively. The median MAEs over current winter and spring wheat harvested areas were less than 0.4 t/ha and 0.3 t/ha for winter and spring wheat emulators, respectively, and the MAEs of the pDSSAT-emulator and LPJmL-emulator were relatively greater. Over current maize harvested areas, the median *R* was greater than ~~0.85~~0.6 and the median of MAE was less than ~~0.4~~0.7 t/ha, except pDSSAT-emulator. The median *R* of emulators developed for rice were greater than ~~0.89~~0.5, and the median MAE were less than ~~0.3~~0.4 t/ha over current rice harvested areas, whereas the performances of pDSSAT-emulator and CARAIB-emulator were relatively worse.

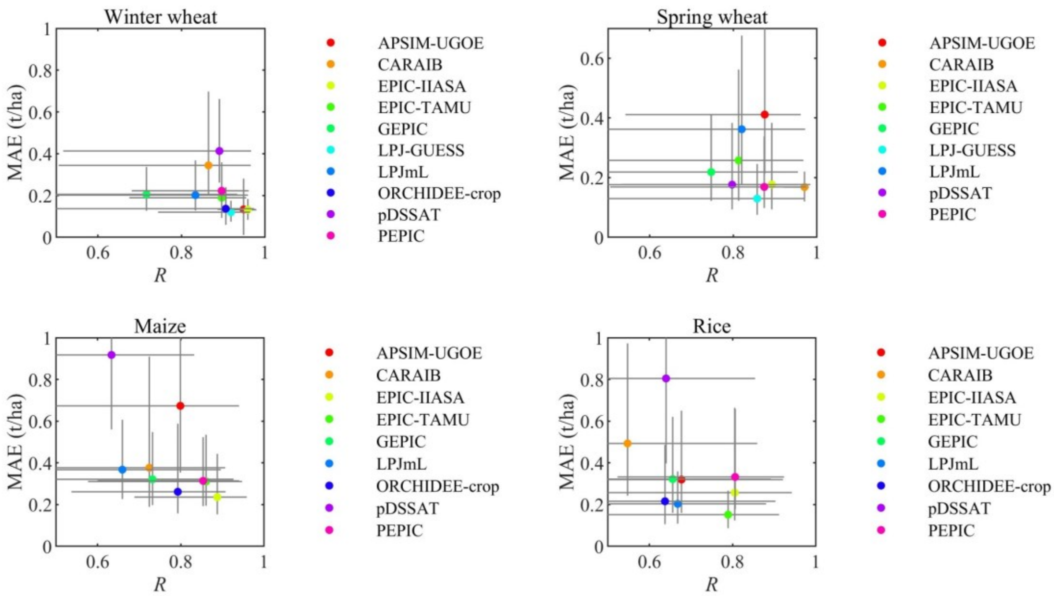

Figure 4 Correlation coefficient (*R*) and mean absolute error (MAE) over current cropland in the baseline (C360T0W0N200). As the ORCHIDEE-crop has not simulated yield under C360T0W0N200 perturbation, we used the C360T0W10N200 as the baseline. The dot denotes the median and the error bar denotes the interquartile range from all grid cells in which the crop is grown according to the SPAM2010 data.

3.2.3 Performance of multiple emulators ensemble at grid scale

The multi-emulators ensemble median was able to reproduce the year-to-year variation of gridded yield over current cropland in the baseline (C360T0W0N200) from 1981 to 2010. The temporal correlation coefficient *R* between GGCM simulated and emulated yield time series over most current harvested areas were greater than ~~0.8~~0.7 (multi-model ensemble median) (**Figure 5**), and the uncertainty (standard deviation) of *R* across emulators was smaller than ~~0.2~~0.3 (Figure ~~S3~~S1). The mean absolute error (MAE) and mean relative error (MRE) over most current harvested areas were mostly smaller than 1 t/a and ~~10~~30%, respectively (Figure ~~S4~~S2). The spatial pattern of MRE for four crops all showed a hotspot of large MRE in the Middle East, and for maize the hotspot of great MRE was also found in the southern China (Figure ~~S4~~S2).

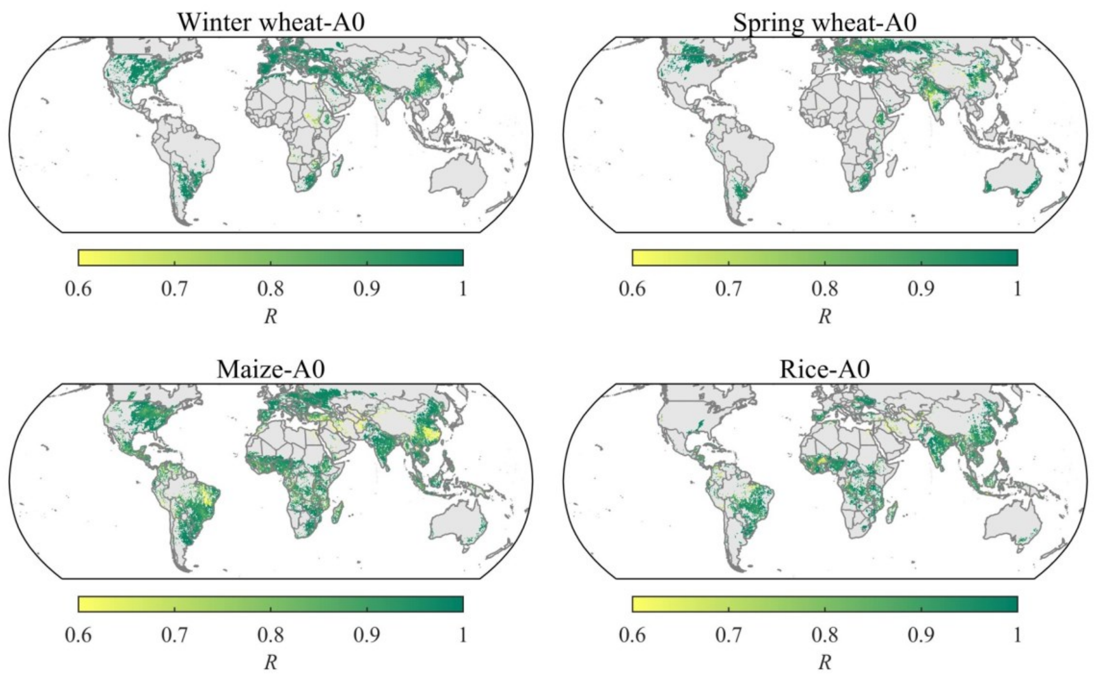

Figure 5 Multi-model ensemble median R in the baseline over current cropland. R: correlation coefficient between simulated and emulated yield time series of each GGCM from 1981 to 2010.

3.3 Performance of emulators to capture the year-to-year variation of GGCM yield in the CTWN cube

3.3.1 Performance of individual emulators at the global scale

The agreement of year-to-year variation of global average yield between simulation and emulation was consistent with changes in CTWN cube over present cropland (**Figure 6**). Under varied CTWN perturbations, the emulator could well reproduce the year-to-year variation of global mean yield from 1981 to 2010. Even when the temperature perturbation reached +6K, the emulator was still able to capture the year-to-year variation of global mean yield. Similarly, when the precipitation was less than baseline by 50%, the year-to-year variation of emulated global mean yield was well matched with those of GGCM simulation. Additionally, the fertilizations of elevated CO_2 concentration and nitrogen application have been well reproduced by emulator. Similar capacity in reproducing the annual global mean yield was also been found in other emulators (~~Figure S5~~Table S1 & Table S2). Even under the concurrent warm and drought condition, i.e. T+6K and W-50%, the year-to-year variation of global mean yield could be well reproduced by emulator (Figure S~~6~~3).

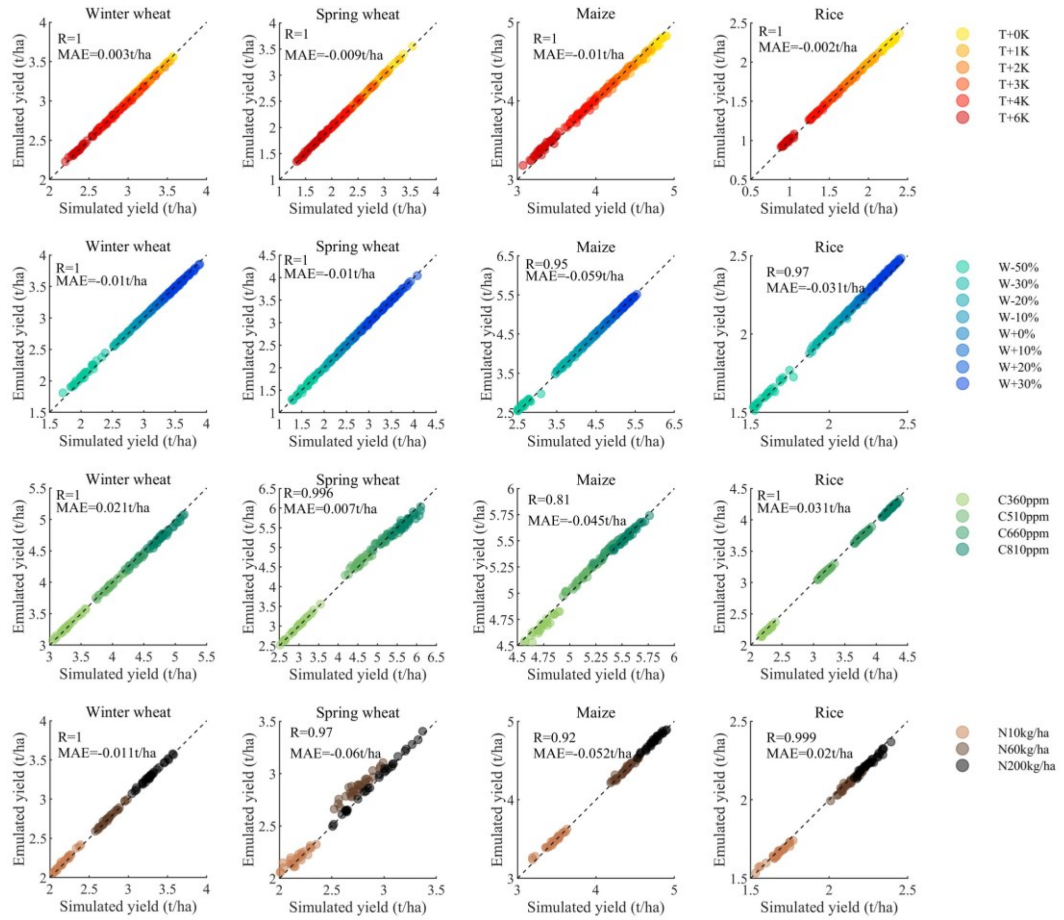

Figure 6 Performance of one exemplary emulator (LPJmL-A0) in reproducing the year to year variation of global mean yield from 1981 to 2010 under varied individual CTWN perturbations. Each point with the same color is yield in one year. The performances of other emulators are similar to LPJmL-A0, which can be referred in the ~~Figure S5 for EPIC-TAMU~~Table S1 and Table S2.

3.3.2 Performance of individual emulators at the grid scale under single perturbation

To illustrate the performance of individual emulators to reproduce annual yield variation, we selected the LPJmL-A0 emulator as an example. The *R*-MAE scatter plots of LPJmL-A0 illustrated the response of gridded accuracy to varied perturbations of CTWN (**Figure 7**). The changes in accuracy of emulators under single CTWN perturbations were small with largest differences in spring wheat for modifications in the CO_2 (C) and nitrogen (N) dimensions~~, and for rice for modifications in the water (W) dimension~~. The overall accuracy could be kept on the high level, with greater *R* and smaller MAE. Under temperature perturbations, the median *R*s of emulators for four crops were greater than ~~0.9~~0.7, and the range of *R*s was smaller than ~~0.03~~0.2. The median MAEs of emulators for four crops were less than ~~0.35~~0.55, and the variation of median MAEs was smaller than ~~0.02~~0.2 from +1 to +6K perturbations. For precipitation perturbations, the median *R*s of emulators for four crops were greater than ~~0.88~~0.85, meanwhile the difference of median *R*s across varied precipitation perturbations was smaller than ~~0.06~~0.1. The median MAEs of emulators for four crops was smaller than ~~0.38~~0.3, and the range of median MAEs variation was as small as ~~0.05~~0.06. The median *R*s and MAEs of emulators for four crops under CO_2 concentration perturbations and nitrogen perturbations were comparable to those under temperature and precipitation perturbations, except for spring wheat. Although the performance of spring wheat emulator under CO_2 and nitrogen perturbations was not as good as other crops, the median *R*s was still greater than 0.75 and the median MAEs were smaller than ~~0.2~~0.6. Similar pattern of other emulators' performances under single perturbations at grid scale are shown in the Table ~~S2~~ S1 and Table ~~S3~~S2.

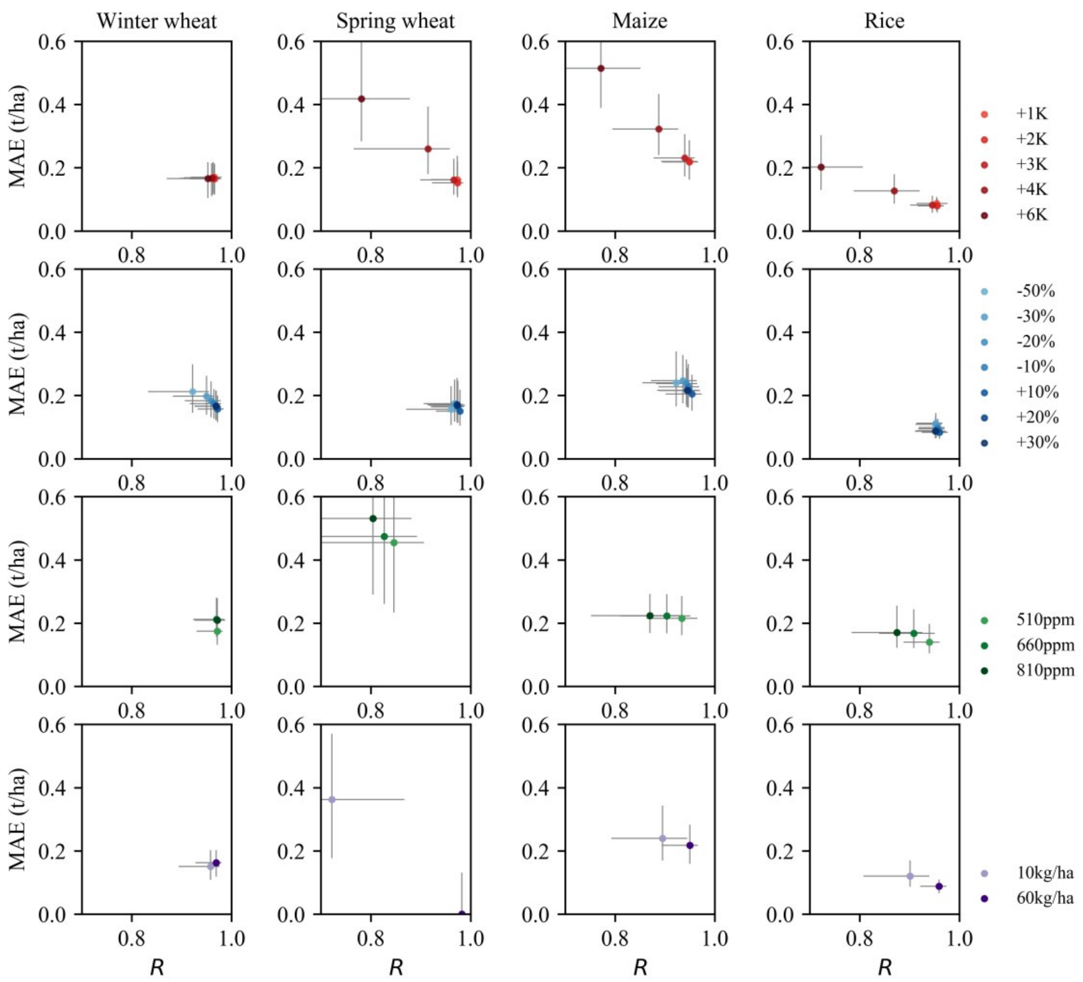

Figure 7 *R*-MAE scatter of the exemplary emulator (LPJmL-A0) under varied single CTWN perturbations. Each dot denotes the median of *R* or MAE over current cropland, the error bar denotes the interquartile range. *R*: correlation coefficient, MAE: mean absolute error. More details of other emulators can refer to Table S~~2~~1 and S~~3~~2.

3.3.3 Performance of multiple emulators ensemble at the grid scale under single perturbation

When looking at the ensemble of multiple emulators, the *R*s and MAEs under CTWN cubes was not divergent obviously (**Figure 8**, **Figure 9**).

Under temperature perturbations, the range of model-ensemble median *R*s across multiple emulators was smaller than ~~0.01~~0.2, and the range of median MAEs was as small as ~~0.03~~0.4t/ha. For precipitation perturbation, the difference in median *R*s was less than 0.03, and the changes in median MAEs was less than ~~0.09~~0.1t/ha. Under the perturbation of CO_2 concentration, the emulators for winter wheat, maize and rice showed a greater median Rs which ranged from ~~0.96~~0.89 to 0.98. The variation of median MAEs was smaller than ~~0.08~~0.09t/ha. The median Rs of emulator for spring wheat, however, tended to decline under 810ppm perturbation substantially and the median MAEs tended to increase simultaneously. Similarly, for nitrogen perturbation, the range of median Rs was less than ~~0.15~~0.27, and the range of median MAEs was smaller than ~~0.1~~0.3t/ha, except for emulators of spring wheat and rice. The declined R and increased MAE were caused by the reduction of valid sample size from the GGCM output yield under CO_2 and nitrogen perturbations (Figure S4~~7~~ & Figure S5~~8~~).

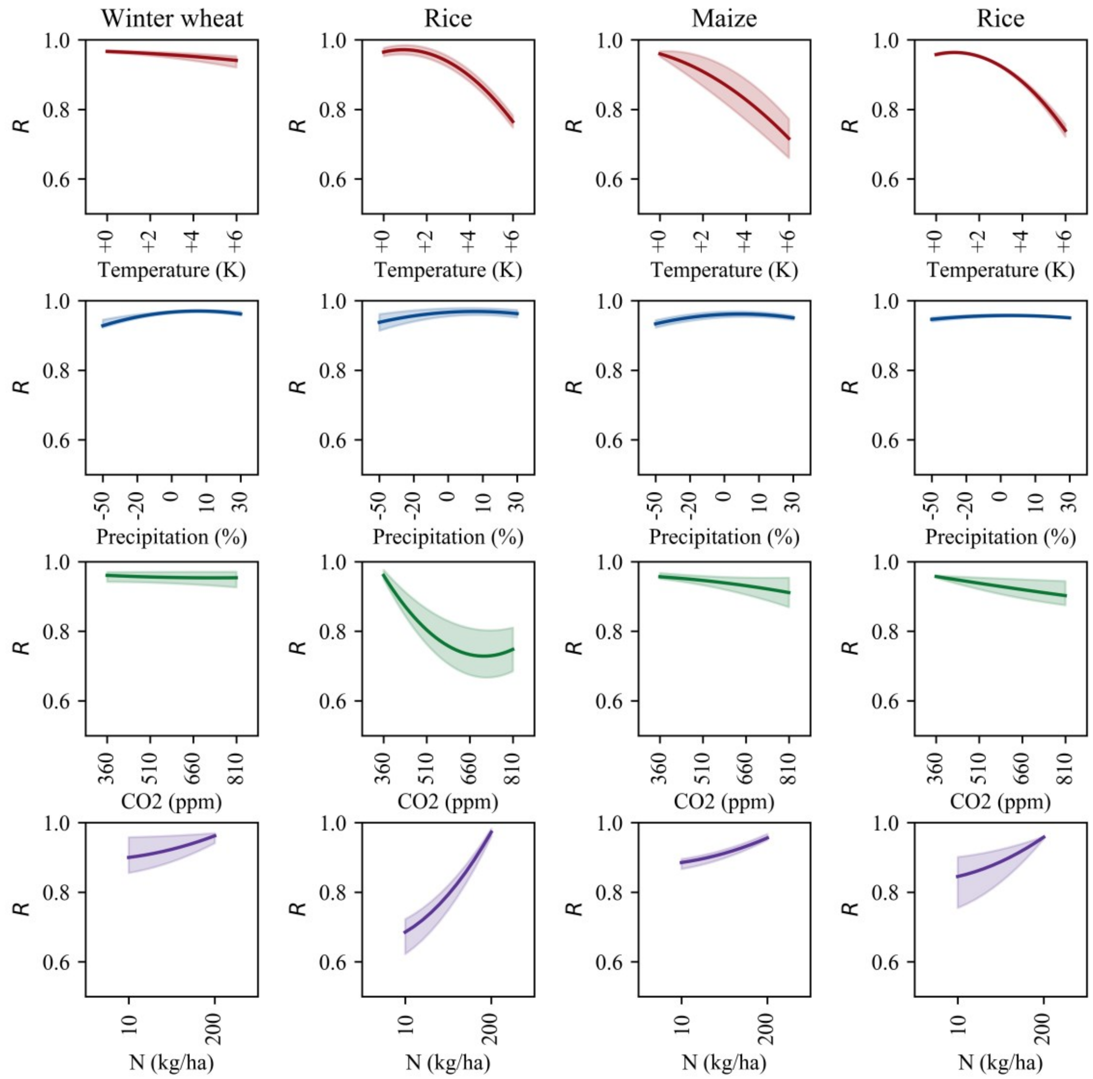

Figure 8 Correlation coefficient (R) of multiple emulators ensemble under varied TW perturbations. The line denotes the median of R over current cropland, and the shaded area denotes the range of median R over current cropland across emulators.

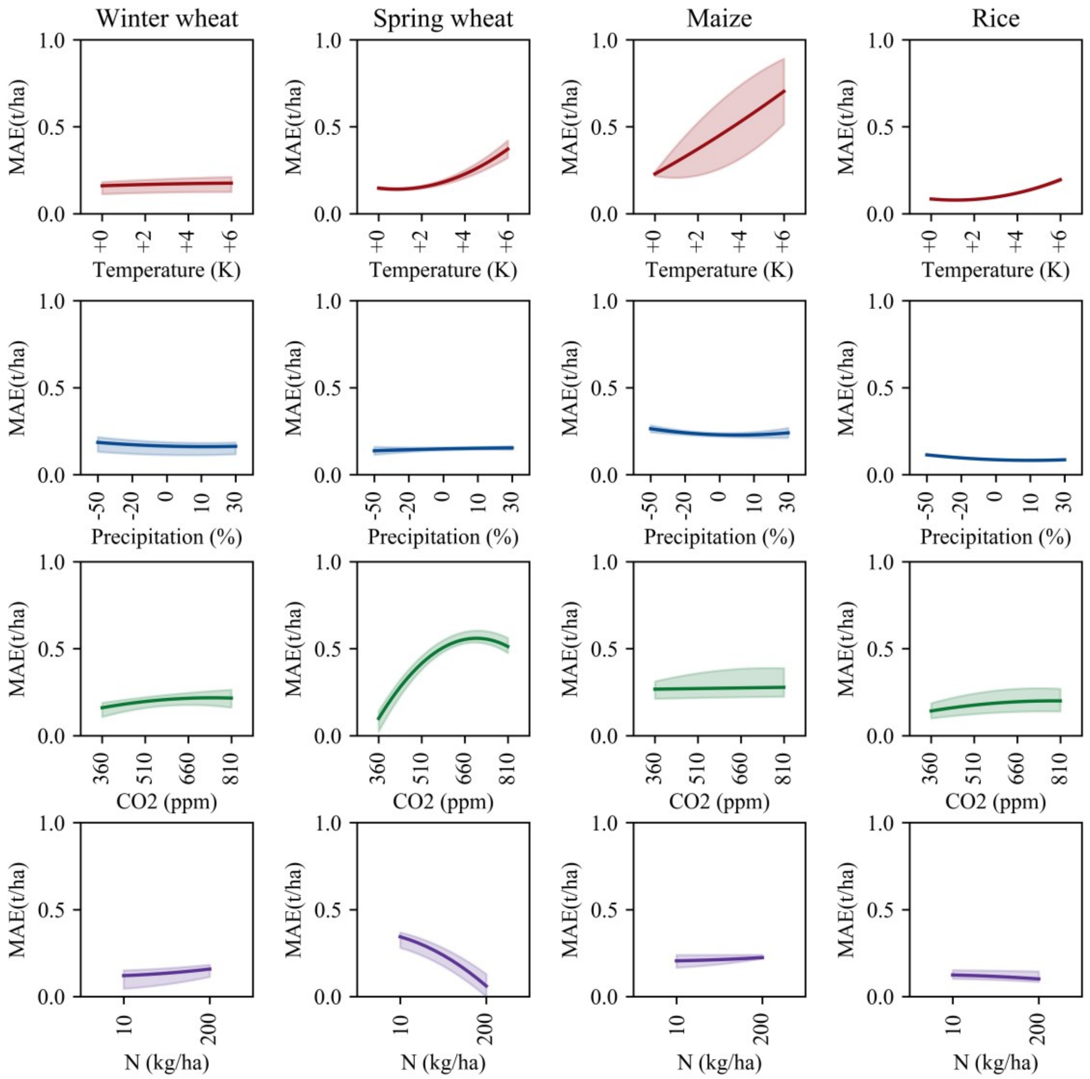

Figure 9 Mean absolute error (MAE) of multiple emulators ensemble under varied CTWN perturbations. The line denotes the median of *R* over current cropland, and the shaded area denotes the range of median *R* over current cropland across emulators.

3.3.4 Performance of multiple emulators at grid scale under dual perturbations

The performance of emulators was influenced by changes in simultaneous perturbations in two different CTWN dimensions (dual perturbations). The emulators performed well over most of current cropland but at extreme increases in T and reductions in W (**Figure 10**), the emulators could represent the GGCMI-simulated year-to-year variation only on substantially smaller shares of the current cropland. The fraction of current areas with *R* greater than 0.8 was the highest in the baseline but decreases under warmer and drier conditions. The fraction reduced to less than ~~60~~40% under compound T+6K and

W-50% perturbation, which illustrated the poor capacity of emulator under compound hot-dry conditions. However, the fraction of harvested areas with MAE smaller than 0.5 t/ha did not vary much across T+W perturbations (**Figure 11**). The performance of emulators under dual perturbations for wheat were better than those for maize and rice. The fraction of maize and rice harvested area with *R* greater than 0.8 was relatively smaller than that of wheat. The maize harvested area with MAE smaller than 0.5 t/ha was smaller than other crops. Among the three GGCMs with full range of CTWN perturbations, the fraction of harvested area with high accuracy for LPJmL-emulator and pDSSAT-emulator was more than EPIC-TAMU-emulator.

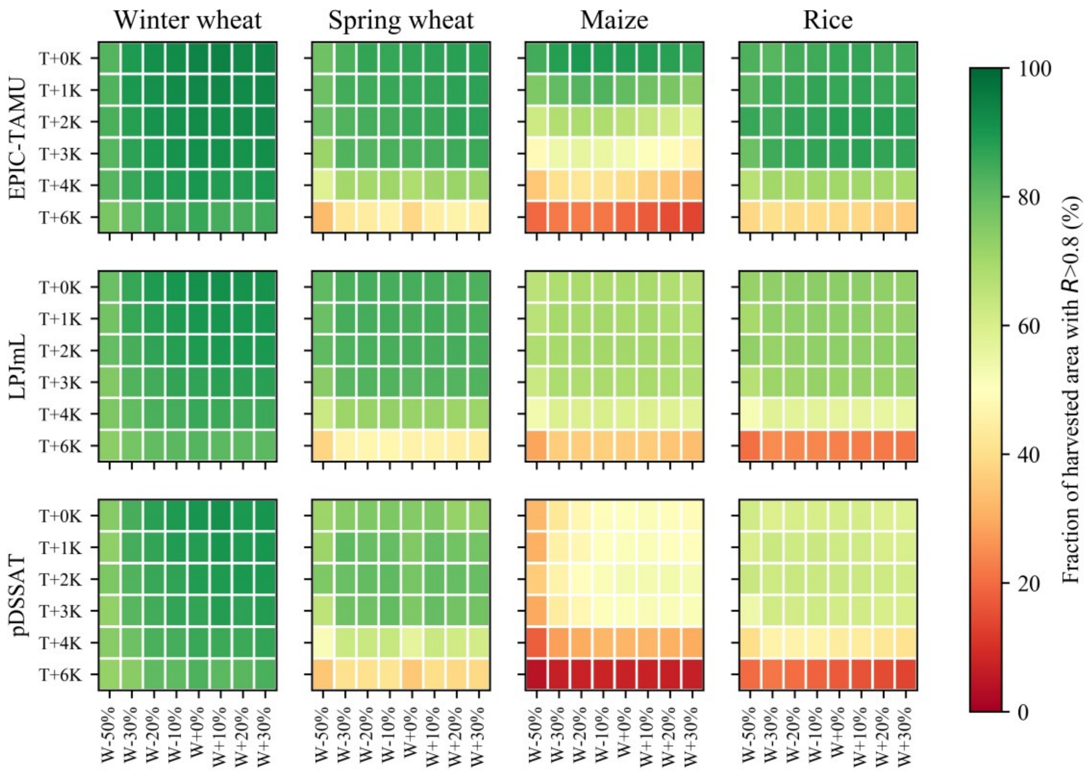

Figure 10 Fraction of harvested areas with high correlation coefficient ($R > 0.8$) under varied T+W perturbations. Example of EPIC-TAMU-A0, LPJmL-A0 and pDSSAT-A0 emulator because only these three GGCMs contain full range of CTWN perturbations for all four crops.

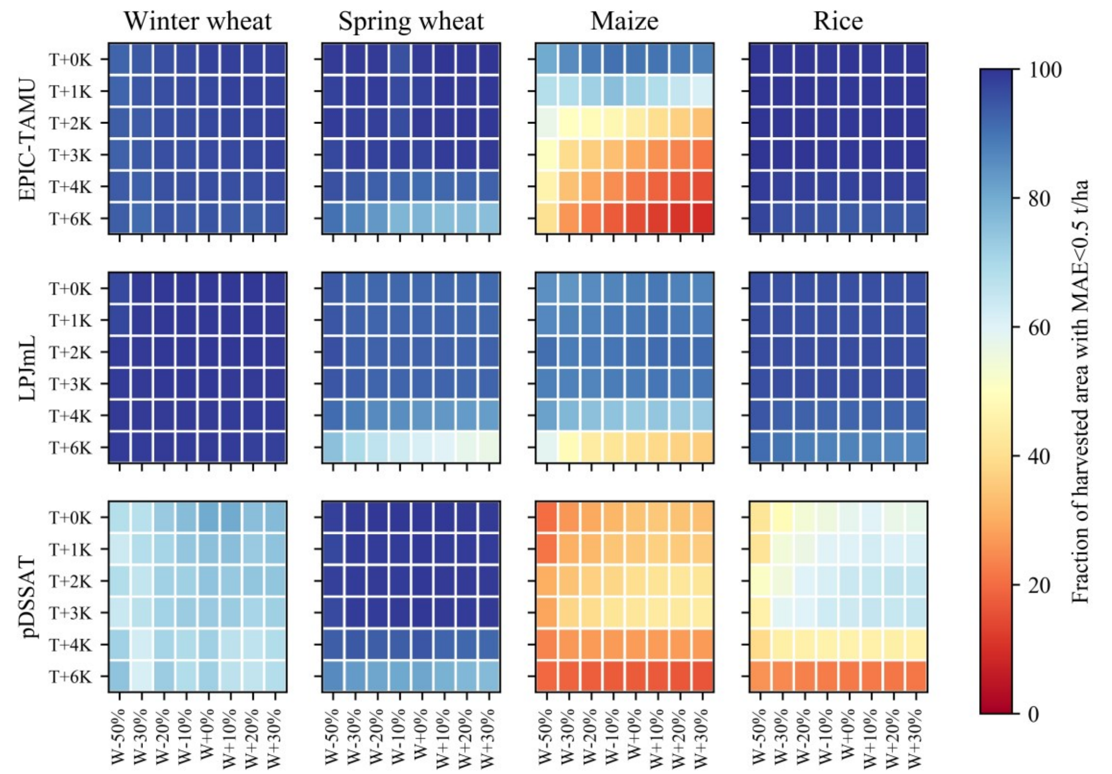

Figure 11 Fraction of harvested areas with low mean absolute error (MAE<0.5 t/ha) under varied T+W perturbations. Example of EPIC-TAMU-A0, LPJmL-A0 and pDSSAT-A0 emulator because these three GGCMs contain full range of CTWN perturbations for all four crops.

4. Discussion

4.1 Emulator trained to capture year-to-year variation in crop yield

Our emulator was designed to reproduce the year-to-year variation of crop yield. Therefore, the annual yield was the target variable in emulator fitting. To capture the year-to-year crop yield variation well, the climatic predictors were divided into growing season average, daily variation and climatic extremes to capture the possible drivers of yield variation. The predictors engineering referred to the existing knowledges compiled into crop models that year-to-year variation of crop yield is associated with growing season temperature and precipitation (Ray et al., 2015), extreme heat (Iizumi and Ramankutty, 2016) and drought (Heinicke et al., 2022). The temperature and precipitation have been confirmed to be the dominant drivers to crop yield variability

(Schauberger et al., 2016). Moreover, the interaction between soil type and climate was considered in our emulator design. Although CO_2 concentration and soil type were not regarded as important contributors to yield variability, their interaction with climate could also influence the yield variability (Kadam et al., 2014). The role of soil type has been uncovered by previous emulator fitted by multivariate regression that the average effect of temperature and precipitation differed greatly depending on soil type (Blanc, 2017). Compared with the emulator designed to reproduce the climatological mean yield, our emulator is more suitable to project the changes in yield variability (Liu et al., 2021b).

We developed the emulators with one statistical relationship for each crop between GGCM simulated yield and predictors for all grids over global lands. Each grid cell represents a sample in the soil-climate-fertilizer continuum, and the training data have no lateral relationships. However, the response of simulated crop yield to climate change was spatially heterogeneous, which mainly depends on the cultivars. Therefore, one statistical relationship between yield and climatic predictors was hard to be fully appropriate for each grid. In response, we used the length of growing season, a representative predictor of cultivar characteristics, to adjust the global statistical relationship to each grid. Therefore, predictors contained both temporal varied and constant variables. The temporal varied predictors were climatic variables which mainly played the role in reproduce the annual yield variation, and the temporal constant predictors were non-climatic variables, like growing season length, delineated the spatial distinction of crop yield response to climate. Compared with region-specific emulator development, combining the temporal varied and constant predictors was more concise and could profit from a broader range of data in the training set.

4.2 Potential application of the well performed emulators in related fields

The good performance over most grid cells indicated the potential capacity of emulators in spatiotemporal downscaling, projecting annual yield in the future and multi-model ensemble simulation.

The emulator could be used to conduct spatiotemporal yield downscaling because the good performance of individual emulator in reproducing the annual crop yield variation in the baseline. As the emulator in this study was developed with a regression-based machine learning technique by using all the grid-year data points, the emulation is not limited to the spatial resolution of the training data. The emulator can be applied to any point with information on the predictors and can produce yield projections is as finely resolved as the forcing input. From the aspect of time series of yield, the raw GGCM data includes empty values ("NaN") in some year-grid cell data points, which may be caused by the lack of regional data for calibration. The vacancy of yield time series in some grids could be imputed by the emulator (Folberth et al., 2019), similar to studies which generated spatiotemporal continuous gridded crop yield data (Chen et al., 2022; Iizumi et al., 2014).

The emulator was able to project the annual yield in the future climate scenarios, which depends on the individual emulator performed well in reproducing annual yield under CTWN cubes. In contrast to many previous emulators developed with historical crop model simulations (Xu et al., 2021), our emulator could reproduce the CO_2 fertilization effect which is an important forcing in future. The recently developed emulator based on GGCMI phase2 simulation under CTWN cubes could only project the climatological-mean yield because the target variable in emulation was the climatological-mean yield (Franke et al., 2020a). In contrast, our emulator can project the annual yield variation and is not constrained by the maximum warming considered in the GGCMI phase2 data set (T+6K), but by the maximum temperature within the training data set (warmest grid cell +6K), so that the applicability is broader (Müller et al., 2021).

It is more efficient to conduct multi-model ensemble simulation with emulators than

GGCMs, as the emulators show good skill in reproducing GGCMs' results and the emulators drastically reduce the computational time and memory requirement and expertise to operate process-based crop models. First, the input of multiple emulators was consistent and compatible but the inputs of raw GGCM were divergent and incompatible because the structure of input data and file format of each GGCM was designed independently. Second, the time-scale of emulator input was monthly or growing seasonal, which was less complex than daily inputs of GGCMs. Apart from the ensemble simulation, the multiple emulators could also be used to explore and disentangle the uncertainty across models.

4.3 Uncertainties

The weaknesses of machine learning algorithm and raw GGCM have brought some uncertainties into the emulators. The uncertainties induced by the machine learning algorithm was as follows:

(1) When the climate factors went beyond the range of training data, the weakness of machine learning in out-of-sample prediction could bring great uncertainty. The emulator inputs should be capped by the range of training data. The limit of our emulator was the warmest grid under +6K perturbation. As there is polar amplification, the strongest warming mostly happens in cooler regions. Thus, the projected temperature exceeding training range would not be widespread over global croplands.

~~(2) The random selection of testing samples in machine learning algorithm failed to warrant independence from training samples when dependence structure exist in the data (Meyer and Pebesma, 2021; Ploton et al., 2020). In our cross-validation, the adjusted R^2s were likely to be overestimated when using the 10-fold cross validation approach with randomly selected trained and validated samples due to the spatial autocorrelated simulated yield. By using the leave-one-out validation approach, the overestimation of adjusted R^2s has been reduced after excluding the spatial~~

~~autocorrelation. Yet, the emulators derived from the 10-fold cross validation and leave-one-out validation approach are not directly comparable in terms of goodness-of-fit statistics due to completely different parameters trained. We then carefully compared the relative feature importance. As shown in Figure S9-S12, relative importance of predictors was consistent across the two validation strategies. That said, the emulator trained and validated by 10-fold cross validation with randomly selected samples can reproduce the climate-yield relationship similar to that derived from the leave-one-out validation approach, in spite of over-estimation of the goodness-of-fit statistics. In model application, we would suggest use the emulators derived from the 10-fold cross validation due to its random sampling to avoid any potential biased estimation. But users still should be cautious when describing the accuracy of the machine learning based emulators. Still, model goodness-of-fits were reasonably good for emulating. Considering the spatial autocorrelation when fitting a machine learning model could provide a more objective understanding of model accuracy.~~

(2~~3~~) Although the emulators could reproduce the GGCM annual yield with high accuracy in most cases, there were cases that the machine learning algorithm did not show good reproduction skill. As the emulated function intended to smooth the response of simulated crop yield to climate, samples at the margins of training data tend to show lower emulator skill. The extreme conditions, i.e. +6K, -50% water, 810ppm, 10kgN/ha, show reduced *R* and increased MAE. Using the emulators to estimate annual crop yield under extreme perturbation conditions should conducted with caution and the additional uncertainty induced by the emulators should be considered in the interpretation of results.

(3~~4~~) Last but not the least, as the emulators are intended as lightweight tools that could replicate the raw GGCMs, their capability in simulating crop yields is limited to the capability of the original GGCMs. This raises the issue that emulators are unlikely to show good performance in simulating crop yield responses to climate extremes, exactly like the raw GGCMs, which have shown poor performance in capturing the yield impact of heatwave and drought (Heinicke et al., 2022), and the lack of negative effect of excessive wetness (Li et al., 2019a). Resolving such a problem requires the improvement of raw GGCMs' capability in simulating yield response to climate extremes, or statistical promotion of the GGCMs' outputs under extreme weather events.

5. Conclusion

In this study, we developed the machine-learning based statistical crop yield emulators to reproduce the year-to-year variation of crop yield to perturbations in CO_2 concentration, temperature, water and nitrogen-application rate from the GGCMI phase 2 archives. To examine the potential value of these emulators, we evaluated the performance of emulators at global and gridded scale under baseline, under single and dual perturbations.

The results indicated that the performance of emulators was good enough to reproduce the year-to-year variation of global average crop yield in the baseline (R > 0.9~~8~~), and the difference of accuracy between individual GGCM emulators were not large. Similarly, under single and dual perturbations, the capacity of emulators in reproducing the year-to-year variation of global mean crop yield was not substantially changed. At gridded level, the performance of emulators over most of the current croplands in the baseline was still good in the sense that R was greater than ~~0.8~~0.6 and MAE was smaller than ~~0.5~~1 t/ha. The performance of individual emulators was consistently good under single CTWN perturbations, without substantial changes in R and MAE. Similarly, the multiple emulators also performed well in reproducing the annual yield under single CTWN perturbations, and the most grid cells across the current cropland showed greater R and smaller MAE under simultaneous perturbations of T and W. The overall good capacity of emulators in reproducing the year-to-year variation of GGCM simulated crop yield indicated the role of emulators in spatiotemporal downscaling, crop yield projection and multi-model ensemble simulation. The emulators were able to boost the ability to assess crop yield failure risk under future climate change and help to better understand food stability and climate risk adaptation.

Code availability

The python function for crop model emulators are available at https://doi.org/10.5281/zenodo.7796686

Author contributions

WL and TY designed the research. WL, TY and CM prepared the manuscript. All authors contributed to editing the manuscript.

Competing interests

Some authors are members of the editorial board of GMD. The peer-review process was guided by an independent editor, and the authors have also no other competing interests to declare.

Acknowledgment

This study was supported by State Key Laboratory of Earth Surface Processes and Resource Ecology of China (2022-ZD-06), the National Natural Science Foundation of China (42171075).

References

Blanc, E. and Sultan, B.: Emulating maize yields from global gridded crop models using statistical estimates, Agric. For. Meteorol., 214–215, 134–147, doi:10.1016/j.agrformet.2015.08.256, 2015.

Blanc, É.: Statistical emulators of maize, rice, soybean and wheat yields from global gridded crop models, Agric. For. Meteorol., 236, 145–161, doi:10.1016/j.agrformet.2016.12.022, 2017.

Blanc, É.: Statistical emulators of irrigated crop yields and irrigation water requirements, Agric. For. Meteorol., 284(January), 107828, doi:10.1016/j.agrformet.2019.107828, 2020.

Campbell, B. M., Vermeulen, S. J., Girvetz, E., Loboguerrero, A. M. and Ramirez-Villegas, J.: Reducing risks to food security from climate change, Glob. Food Sec., 11, 34–43, doi:10.1016/j.gfs.2016.06.002, 2016.

Chen, S., Liu, W., Feng, P., Ye, T., Ma, Y. and Zhang, Z.: Improving Spatial Disaggregation of Crop Yield by Incorporating Machine Learning with Multisource Data: A Case Study of Chinese Maize Yield, Remote Sens., 14(10), doi:10.3390/rs14102340, 2022.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, ACM, New York, NY, USA., 2016.

Elliott, J., Müller, C., Deryng, D., Chryssanthacopoulos, J., Boote, K. J., Büchner, M., Foster, I., Glotter, M., Heinke, J., Iizumi, T., Izaurralde, R. C., Mueller, N. D., Ray, D. K., Rosenzweig, C., Ruane, A. C. and Sheffield, J.: The Global Gridded Crop Model Intercomparison: data and modeling protocols for Phase 1 (v1.0), Geosci. Model Dev., 8(2), 261–277, doi:10.5194/gmd-8-261-2015, 2015.

Folberth, C., Baklanov, A., Balkovič, J., Skalský, R., Khabarov, N. and Obersteiner, M.: Spatio-temporal downscaling of gridded crop model yield estimates based on machine learning, Agric. For. Meteorol., 264(May 2018), 1–15, doi:10.1016/j.agrformet.2018.09.021, 2019.

Franke, J. A., Müller, C., Elliott, J., Ruane, A. C., Jägermeyr, J., Balkovic, J., Ciais, P., Dury, M., Falloon, P. D., Folberth, C., François, L., Hank, T., Hoffmann, M., Izaurralde, R. C., Jacquemin, I., Jones, C., Khabarov, N., Koch, M., Li, M., Liu, W., Olin, S., Phillips, M., Pugh, T. A. M., Reddy, A., Wang, X., Williams, K., Zabel, F. and Moyer, E. J.: The GGCMI Phase 2 emulator: Global gridded crop model response to changes in CO2, temperature, water, and nitrogen (protocol version 1.0), Geosci. Model Dev., 13(5), 2315–2336, doi:10.5194/gmd-13-2315-2020, 2020a.

Franke, J. A., Müller, C., Elliott, J., Ruane, A. C., Jägermeyr, J., Balkovic, J., Ciais, P., Dury, M., Falloon, P. D., Folberth, C., François, L., Hank, T., Hoffmann, M., Izaurralde, R. C., Jacquemin, I., Jones, C., Khabarov, N., Koch, M., Li, M., Liu, W., Olin, S., Phillips, M., Pugh, T. A. M., Reddy, A., Wang, X., Williams, K., Zabel, F. and Moyer, E. J.: The GGCMI Phase 2 experiment: Global gridded crop model simulations under uniform changes in CO2, temperature, water, and nitrogen levels (protocol version 1.0), Geosci. Model Dev., 13(5), 2315–2336, doi:10.5194/gmd-13-2315-2020, 2020b.

Frieler, K., Schauberger, B., Arneth, A., Balkoviˇ, J., Elliott, J., Folberth, C., Deryng, D., Müller, C., Olin, S., Pugh, T. A. M., Schaphoff, S., Schewe, J., Schmid, E., Warszawski, L. and Levermann, A.: Understanding the weather signal in national crop-yield variability Earth ’ s Future, Earth’s Futur., 5, 605–616, doi:10.1002/eft2.217, 2017.

Fronzek, S., Pirttioja, N., Carter, T. R., Bindi, M., Hoffmann, H., Palosuo, T., Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M. F., Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P., Jacquemin, I., Kersebaum, K. C., Kollas, C., Krzyszczak, J., Lorite, I. J., Minet, J., Minguez, M. I., Montesino, M., Moriondo, M., Müller, C., Nendel, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A. C., Ruget, F., Sanna, M., Semenov, M. A., Slawinski, C., Stratonovitch, P., Supit, I., Waha, K., Wang, E., Wu, L., Zhao, Z. and Rötter, R. P.: Classifying multi-model wheat yield impact response surfaces showing sensitivity to temperature and precipitation change, Agric. Syst., 159(June 2017), 209–224, doi:10.1016/j.agsy.2017.08.004, 2018.

Hasegawa, T., Sakurai, G., Fujimori, S., Takahashi, K., Hijioka, Y. and Masui, T.: Extreme climate events increase risk of global food insecurity and adaptation needs, Nat. Food, 2(8), 587–595, doi:10.1038/s43016-021-00335-4, 2021.

Heinicke, S., Frieler, K., Jägermeyr, J. and Mengel, M.: Global gridded crop models underestimate yield responses to droughts and heatwaves, Environ. Res. Lett., 0–68 [online] Available from: https://iopscience.iop.org/article/10.1088/1748-9326/ac592e, 2022.

Iizumi, T. and Ramankutty, N.: Changes in yield variability of major crops for 1981-2010 explained by climate change, Environ. Res. Lett., 11(3), 34003, doi:10.1088/1748-9326/11/3/034003, 2016.

Iizumi, T., Yokozawa, M., Sakurai, G., Travasso, M. I., Romanenkov, V., Oettli, P. and Newby, T.: Historical changes in global yields : major cereal and legume crops from 1982 to 2006, , 346–357, doi:10.1111/geb.12120, 2014.

Jägermeyr, J., Robock, A., Elliott, J., Muller, C., Xia, L., Khabarov, N., Folberth, C., Schmid, E., Liu, W., Zabel, F., Rabin, S. S., Puma, M. J., Heslin, A., Franke, J., Foster, I., Asseng, S., Bardeen, C. G., Toon, O. B. and Rosenzweig, C.: A regional nuclear conflict would compromise global food security, Proc. Natl. Acad. Sci. U. S. A., 117(13), 7071–7081, doi:10.1073/pnas.1919049117, 2020.

Jägermeyr, J., Müller, C., Ruane, A., Elliott, J., Balkovic, J., Castillo, O., Faye, B., Foster, I., Folberth, C., Franke, J., Fuchs, K., Guarin, J., Heinke, J., Hoogenboom, G., Iizumi, T., Jain, A. ., Kelly, D., Khabarov, N., Lange, S., Lin, T., Liu, W., Mialyk, O., Minol, S. and Rosenzweig, C.: Climate change signal in global agriculture emerges earlier in new generation of climate and crop models, Nat. Food (in Revis., 2021.

Janssens, C., Havlík, P., Krisztin, T., Baker, J., Frank, S., Hasegawa, T., Leclère, D., Ohrel, S., Ragnauth, S., Schmid, E., Valin, H., Van Lipzig, N. and Maertens, M.: Global hunger and climate change adaptation through international trade, Nat. Clim. Chang., 10(9), 829–835, doi:10.1038/s41558-020-0847-4, 2020.

Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., Godfray, H. C. J., Herrero, M., Howitt, R. E., Janssen, S., Keating, B. A., Munoz-Carpena, R., Porter, C. H., Rosenzweig, C. and Wheeler, T. R.: Brief history of agricultural systems modeling, Agric. Syst., 155, 240–254, doi:10.1016/j.agsy.2016.05.014, 2017.

Kadam, N. N., Xiao, G., Melgar, R. J., Bahuguna, R. N., Quinones, C., Tamilselvan, A., Prasad, P. V. V and Jagadish, K. S. V: Chapter Three - Agronomic and Physiological Responses to High Temperature, Drought, and Elevated CO2 Interactions in Cereals, vol. 127, edited by D. B. T.-A. in A. Sparks, pp. 111–156, Academic Press., 2014.

Kinnunen, P., Guillaume, J. H. A., Taka, M., D’Odorico, P., Siebert, S., Puma, M. J., Jalava, M. and Kummu, M.: Local food crop production can fulfil demand for less than one-third of the population, Nat. Food, 1(4), 229–237, doi:10.1038/s43016-020-0060-7, 2020.

Li, Y., Guan, K., Schnitkey, G. D., DeLucia, E. and Peng, B.: Excessive rainfall leads to maize yield loss of a comparable magnitude to extreme drought in the United States, Glob. Chang. Biol., 25(7), 2325–2337, doi:10.1111/gcb.14628, 2019a.

Li, Y., Guan, K., Yu, A., Peng, B., Zhao, L., Li, B. and Peng, J.: Toward building a transparent statistical model for improving crop yield prediction: Modeling rainfed corn in the U.S, F. Crop. Res., 234(January), 55–65, doi:10.1016/j.fcr.2019.02.005, 2019b.

Liu, W., Ye, T. and Shi, P.: Decreasing wheat yield stability on the North China Plain: Relative contributions from climate change in mean and variability, Int. J. Climatol., 41(S1), E2820–E2833, doi:10.1002/joc.6882, 2021a.

Liu, W., Ye, T., Jägermeyr, J., Müller, C., Chen, S., Liu, X. and Shi, P.: Future climate change significantly alters interannual wheat yield variability over half of harvested areas, Environ. Res. Lett., 16(9), 094045, doi:10.1088/1748-9326/ac1fbb, 2021b.

Liu, W., Li, Z., Li, Y., Ye, T., Chen, S. and Liu, Y.: Heterogeneous impacts of excessive wetness on maize yields in China: Evidence from statistical yields and process-based crop models, Agric. For. Meteorol., 327(August), 109205, doi:10.1016/j.agrformet.2022.109205, 2022.

Lobell, D. B., Sibley, A. and Ivan Ortiz-Monasterio, J.: Extreme heat effects on wheat senescence in India, Nat. Clim. Chang., 2(3), 186–189, doi:10.1038/nclimate1356, 2012.

Makowski, D., Asseng, S., Ewert, F., Bassu, S., Durand, J. L., Li, T., Martre, P., Adam, M., Aggarwal, P. K., Angulo, C., Baron, C., Basso, B., Bertuzzi, P., Biernath, C., Boogaard, H., Boote, K. J., Bouman, B., Bregaglio, S., Brisson, N., Buis, S., Cammarano, D., Challinor, A. J., Confalonieri, R., Conijn, J. G., Corbeels, M., Deryng, D., De Sanctis, G., Doltra, J., Fumoto, T., Gaydon, D., Gayler, S., Goldberg, R., Grant, R. F., Grassini, P., Hatfield, J. L., Hasegawa, T., Heng, L., Hoek, S., Hooker, J., Hunt, L. A., Ingwersen, J., Izaurralde, R. C., Jongschaap, R. E. E., Jones, J. W., Kemanian, R. A., Kersebaum, K. C., Kim, S. H., Lizaso, J., Marcaida, M., Müller, C., Nakagawa, H., Naresh Kumar, S., Nendel, C., O'Leary, G. J., Olesen, J. E., Oriol, P., Osborne, T. M., Palosuo, T., Pravia, M. V., Priesack, E., Ripoche, D., Rosenzweig, C., Ruane, A. C., Ruget, F., Sau, F., Semenov, M. A., Shcherbak, I., Singh, B., Singh, U., Soo, H. K., Steduto, P., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tang, L., Tao, F., Teixeira, E. I., Thorburn, P., Timlin, D., Travasso, M., Rötter, R. P., Waha, K., Wallach, D., White, J. W., Wilkens, P., Williams, J. R., Wolf, J., Yin, X., Yoshida, H., Zhang, Z. and Zhu, Y.: A statistical analysis of three ensembles of crop model responses to temperature and CO2 concentration, Agric. For. Meteorol., 214–215, 483–493, doi:10.1016/j.agrformet.2015.09.013, 2015.

Meyer, H. and Pebesma, E.: Predicting into unknown space? Estimating the area of applicability of spatial prediction models, Methods Ecol. Evol., 12(9), 1620–1633, doi:10.1111/2041-210X.13650, 2021.

Mistry, M. N., Sue Wing, I. and De Cian, E.: Simulated vs. empirical weather responsiveness of crop yields: US evidence and implications for the agricultural impacts of climate change, Environ. Res. Lett., 12(7), doi:10.1088/1748-9326/aa788c, 2017.

Müller, C., Franke, J., Jägermeyr, J., Ruane, A. C., Elliott, J., Moyer, E., Heinke, J., Falloon, P., Folberth, C., Francois, L., Hank, T., Izaurralde, R. C., Jacquemin, I., Liu, W., Olin, S., Pugh, T., Williams, K. E. and Zabel, F.: Exploring uncertainties in global crop yield projections in a large ensemble of crop models and CMIP5 and CMIP6 climate scenarios, Environ. Res. Lett., doi:10.1088/1748-9326/abd8fc, 2021.

Nachtergaele, F., Velthuizen, H. Van, Verelst, L., Batjes, N., Dijkshoorn, K., Engelen, V. Van, Fischer, G., Jones, A., Montanarella, L., Petri, M., Prieler, S., Teixeira, E., Wiberg, D. and Shi, X.: Harmonized World Soil Database (version 1), Soil Sci., p.38, doi:3123, 2009.

Ostberg, S., Schewe, J., Childers, K. and Frieler, K.: Changes in crop yields and their variability at different levels of global warming, Earth Syst. Dyn., 9(2), 479–496, doi:10.5194/esd-9-479-2018, 2018.

Pirttioja, N., Carter, T. R., Fronzek, S., Bindi, M., Hoffmann, H., Palosuo, T., Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M. F., Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P., Jacquemin, I., Kersebaum, K. C., Kollas, C., Krzyszczak, J., Lorite, I. J., Minet, J., Minguez, M. I., Montesino, M., Moriondo, M., Müller, C., Nendel, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A. C., Ruget, F., Sanna, M., Semenov, M. A., Slawinski, C., Stratonovitch, P., Supit, I., Waha, K., Wang, E., Wu, L., Zhao, Z. and Rötter, R. P.: Temperature and precipitation effects on wheat yield across a European transect: A crop model ensemble analysis using impact response surfaces, Clim. Res., 65, 87–105, doi:10.3354/cr01322, 2015.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S. and Pélissier, R.: Spatial validation reveals poor predictive performance of large-scale ecological mapping models, Nat. Commun., 11(1), 1–11, doi:10.1038/s41467-020-18321-y, 2020.

Portmann, F. T., Siebert, S. and Döll, P.: MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling, Global Biogeochem. Cycles, 24(1), doi:10.1029/2008GB003435, 2010.

Raimondo, M., Nazzaro, C., Marotta, G. and Caracciolo, F.: Land degradation and climate change: Global impact on wheat yields, L. Degrad. Dev., 32(1), 387–398, doi:10.1002/ldr.3699, 2021.

Ray, D. K., Gerber, J. S., Macdonald, G. K. and West, P. C.: Climate variation explains a third of global crop yield variability, Nat. Commun., 6, 1–9, doi:10.1038/ncomms6989, 2015.

Ruane, A. C., Goldberg, R. and Chryssanthacopoulos, J.: Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation, Agric. For. Meteorol., 200, 233–248, doi:10.1016/j.agrformet.2014.09.016, 2015.

Sacks, W. J., Deryng, D., Foley, J. A. and Ramankutty, N.: Crop planting dates: an analysis of global patterns, Glob. Ecol. Biogeogr., 19(5), 607–620, doi:10.1111/j.1466-8238.2010.00551.x, 2010.

Schauberger, B., Rolinski, S. and Müller, C.: A network-based approach for semi-quantitative knowledge mining and its application to yield variability, Environ. Res. Lett., 11(12), doi:10.1088/1748-9326/11/12/123001, 2016.

Shahhosseini, M., Martinez-Feria, R. A., Hu, G. and Archontoulis, S. V.: Maize yield and nitrate loss prediction with machine learning algorithms, Environ. Res. Lett., 14(12), 124026, doi:10.1088/1748-9326/ab5268, 2019.

Sternberg, T.: Regional drought has a global impact, Nature, 472(7342), 169–169, doi:10.1038/472169d, 2011.

Tartarini, S., Vesely, F., Movedi, E., Radegonda, L., Pietrasanta, A., Recchi, G. and Confalonieri, R.: Biophysical models and meta-modelling to reduce the basis risk in index-based insurance: A case study on winter cereals in Italy, Agric. For. Meteorol., 300, 108320, doi:https://doi.org/10.1016/j.agrformet.2021.108320, 2021.

Troy, T. J., Kipgen, C. and Pal, I.: The impact of climate extremes and irrigation on US crop yields, Environ. Res. Lett., 10(5), 1–10, doi:10.1088/1748-9326/10/5/054013, 2015.

Xu, H., Zhang, X., Ye, Z., Jiang, L., Qiu, X., Tian, Y., Zhu, Y. and Cao, W.: Machine learning approaches can reduce environmental data requirements for regional yield potential simulation, Eur. J. Agron., 129(August 2020), doi:10.1016/j.eja.2021.126335, 2021.

Zhu, X. and Troy, T. J.: Agriculturally Relevant Climate Extremes and Their Trends in the World's Major Growing Regions, Earth's Futur., 6(4), 656–672, doi:10.1002/2017EF000687, 2018.