European Journal of Agronomy 149 (2023) 126917

Contents lists available at ScienceDirect

European Journal of Agronomy

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

Integrating machine learning and environmental variables to constrain uncertainty in crop yield change projections under climate change

Linchao Li[a,b,1], Yan Zhang[a,b,1], Bin Wang[b,c], Puyu Feng[d], Qinsi He[c,e], Yu Shi[a,b], Ke Liu[f], Matthew Tom Harrison[f], De Li Liu[c,g,*], Ning Yao[h], Yi Li[h], Jianqiang He[h], Hao Feng[a,h], Kadambot H.M. Siddique[i], Qiang Yu[a,b,**]

[a] College of Natural Resources and Environment, Northwest A&F University, Yangling 712100, China
[b] State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Soil and Water Conservation, Northwest A&F University, Yangling 712100, China
[c] NSW Department of Primary Industries, Wagga Wagga Agricultural Institute, Wagga Wagga, NSW 2650, Australia
[d] College of Land Science and Technology, China Agricultural University, Beijing 100193, China
[e] School of Life Sciences, Faculty of Science, University of Technology Sydney, P.O. Box 123, Broadway, NSW 2007, Australia
[f] Tasmanian Institute of Agriculture, University of Tasmania, Newnham Drive, Launceston, Tasmania 7248, Australia
[g] Climate Change Research Centre, University of New South Wales, Sydney, NSW 2052, Australia
[h] College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling 712100, China
[i] The UWA Institute of Agriculture and UWA School of Agriculture and Environment, The University of Western Australia, Perth, WA 6001, Australia

ARTICLE INFO

Keywords:
Yield projections
Crop modelling
Extreme climate event
Machine learning model
Model uncertainty
Climate change impact

ABSTRACT

Robust crop yield projections under future climates are fundamental prerequisites for reliable policy formation. Both process-based crop models and statistical models are commonly used for this purpose. Process-based models tend to simplify processes, minimize the effects of extreme events, and ignore biotic pressures, while statistical models cannot deterministically capture intricate biological and physiological processes underpinning crop growth. We attempted to integrate and overcome shortcomings in both modelling frameworks by integrating the dynamic linear model (DLM) and random forest machine learning model (RF) with nine global gridded crop models (GGCM), respectively, in order to improve projections and reduce uncertainties of maize (*Zea mays* L.) and soybean (*Glycine max* [L.] Merrill) yield projections. Our results demonstrated substantial improvements in model performance accuracy by using RF in concert with GGCM across China's maize and soybean belt. This improvement surpasses that achieved using DLM. For maize, the GGCM+RF models increased the r values from 0.15 to 0.61–0.64–0.77 and decreased nRMSE from approximately 0.20 to 0.50–0.13–0.17 compared with using GGCM alone. For soybean, the models increased r from 0.37 to 0.70–0.54–0.70 and decreased nRMSE from 0.17 to 0.35–0.17–0.20 compared with using GGCM alone. The main factors influencing maize yield changes included chilling days (CD), crop pests and diseases (CPDs), and drought, while for soybean the primary influencing factors included CPD, tropical days (based on exceeding a maximum temperature), and drought. Our approach decreased uncertainties by 33–78% for maize and by 56–68% for soybean. The main source of uncertainty for GGCM was the crop model. For GGCM+RF, the main source of uncertainty for the 2040–2069 period was the global climate model, while the main source of uncertainty for the 2070–2099 period was the climate scenario. Our results provide a novel, robust, and pragmatic framework to constrain uncertainties in order to accurately assess the impact of future climate change on crop yields. These results could be used to interpret future ensemble studies by accounting for uncertainty in crop and climate models, as well as to assess future emissions scenarios.

* Corresponding author at: NSW Department of Primary Industries, Wagga Wagga Agricultural Institute, Wagga Wagga, NSW 2650, Australia.
** Corresponding author at: College of Natural Resources and Environment, Northwest A&F University, Yangling 712100, China.
E-mail addresses: de.li.liu@dpi.nsw.gov.au (D.L. Liu), yuq@nwafu.edu.cn (Q. Yu).
[1] These authors contributed equally.

https://doi.org/10.1016/j.eja.2023.126917
Received 29 April 2023; Received in revised form 20 June 2023; Accepted 16 July 2023
Available online 20 July 2023
1161-0301/

1. Introduction

Climate change has led to a declining number of extreme cold events (Harrison, 2021) together with an increase in the number of heatwaves (Langworthy et al., 2018; Liu et al., 2016), droughts (Lobell et al., 2014), and excessive rainfall (Li et al., 2019b; Liu et al., 2023) that can significantly impact global food production (Mbow et al., 2019). Maize (*Zea mays* L.) and soybean (*Glycine max* [L.] Merrill) are two of the world's most important staple food crops, accounting for over 50% of global grain production (Liu et al., 2021b). In China, these crops contribute about 30% and 6% to global maize and soybean production, respectively (FAO, 2020). Crop models are commonly used to develop effective adaptation strategies for maintaining stable grain production under future climate change (Huang et al., 2020). However, these models have limitations and uncertainties due to the model structure, parameters, and global climate model (GCM) inputs, particularly under extreme events (Asseng et al., 2013; Li et al., 2019b; Müller et al., 2021; Rosenzweig et al., 2014). For example, Müller et al. (2021) demonstrated substantial uncertainties in global crop yield projections based on a large model ensemble, hindering our ability to provide reliable crop yield projections and to develop effective adaptation strategies (Huang et al., 2022).

Extreme climate events (ECEs) can affect crop growth and yields by subjecting crops to various stresses (Asseng et al., 2014; Lesk et al., 2016; Liu et al., 2023), with some process-based crop models oversimplifying these complex processes (Feng et al., 2019). For instance, the Environmental Policy Integrated Climate (EPIC) model fails to fully capture yield losses under extremely wet conditions (Balkovič et al., 2013), and the JULES model underestimates crop yields in arid irrigated regions, but overestimates yields in tropical regions (Osborne et al., 2015). Similarly, the yield gaps identified in global gridded crop models (GGCM) suggest that global process-based crop models tend to underestimate the magnitude of crop yield losses caused by extreme heatwaves and excessive rainfall (Heinicke et al., 2022; Li et al., 2019b; Liu et al., 2020), leading some authors to suggest that the multi-model median is the most adequate descriptor of model performance (Sandor et al., 2020). Moreover, crop pests and diseases (CPDs) are not considered in most process-based crop models (Jägermeyr et al., 2021; Rosenzweig et al., 2014; Xiong et al., 2019), and this omission may lead to insufficient accounting of additional yield losses (Deutsch et al., 2018). The mechanisms by which CPDs impact crops vary widely, and can potentially impact any combination of biomass, leaf area, light interception, and/or photosynthetic rates of affected plants, making the diversity of such processes difficult to capture in crop models (Bondad et al., 2023).

Compared with process-based crop models, statistical-based models, such as machine learning algorithms, are able to capture potential non-linear relationships between extreme climate events and crop yields without requiring a complex set of parameters and a deep understanding of physical processes (Feng et al., 2019; Li et al., 2022). Machine learning algorithms can incorporate indices such as ECEs and CPDs, as well as genotype by management by environment interactions (Ibrahim et al., 2019), and may require less parameterization than process-based models (Harrison et al., 2019). Several studies have used statistical-based crop models such as machine learning or deep learning algorithms to predict crop yields with acceptable performance (Cao et al., 2021a; Jiang et al., 2020; Li et al., 2021a). However, these models may not be able to capture the biological and physiological processes that influence crop growth, such as the CO_2 fertilization effect. Rather, they only capture the statistical yield outcome. Moreover, the results obtained from statistical-based models may lack interpretation and be less transparent than those obtained from process-based models (Feng et al., 2020).

Numerous studies have developed hybrid models to combine the strengths of process-based and statistical-based crop models (Tao et al., 2022). For example, Feng et al. (2019) integrated APSIM model output with ECE using machine learning to improve modeling accuracy in the New South Wales wheat belt of Australia, increasing R^2 by 0.23. Similarly, Li et al. (2021b) incorporated machine learning with the MCWLA-Wheat model to develop a drought risk assessment system, reducing the root mean squared error (RMSE) from 530 to 365 kg/ha. Such studies only focus on the integration of statistical models into one specific crop model. Because hybrid models can consider the effect of ECEs using large amounts of experimental data, they may also reduce redundancy and further constrain uncertainty in crop yield projections from multi-model ensembles. However, most previous studies have overlooked the importance of reducing the overall uncertainties and compared the sources of uncertainties as projected by crop models and hybrid models. Addressing this gap could enhance the robustness of crop yield projections and improve the understanding of associated uncertainties.

Because different models have varied responses to climate factors, recent studies have explored methods to reduce the uncertainty of crop yield projections under climate change, such as improving temperature response functions (Wang et al., 2017) and emergent constraint methods (Wang et al., 2020c; Yin and Leng, 2022; Zhao et al., 2016b). For instance, Wang et al. (2020c) employed an emergent constraint approach to constrain crop yield responses to temperature in GGCM, utilizing data from 48 field warming experiment sites for wheat, maize, rice, and soybean. This approach reduced uncertainties by 12–54% across the four crops. However, as the relationship between crop yield and ECEs is usually non-linear, machine learning may be better suited to capture the potential non-linear response (Lobell et al., 2011). In addition, other factors that could impact crop yields, such as excessive rainfall (Lesk et al., 2020), drought stress (Wang et al., 2022a), and CPDs (Wang et al., 2021a), should be considered when constraining uncertainty in crop yield projections.

Our study presents a novel method that combines nine GGCMs with a machine learning algorithm to enhance reliability and reduce uncertainty in crop yield projections for maize and soybean. Our objectives were to (1) improve model performance by reproducing observed maize and soybean yields from historical data, (2) investigate the non-linear response of crop yields to ECE and CPD risk, (3) compare yield projections from the GGCM alone against yield projections from hybrid models under future climate scenarios, and (4) analyze the sources of uncertainty in yield projections. Achieving these objectives will provide valuable insights into constraining model uncertainty to assist policymakers and stakeholders in making informed decisions for sustainable agriculture.

2. Data and methodology

2.1. Study area

China's diverse climate zones can impact crop yields differently (Piao et al., 2010; Tao et al., 2008). For example, solar radiation and precipitation changes have increased wheat yields by 1–13% in northern China and decreased wheat yields in southern China by 1.2–10.2% (Tao et al., 2014). It is clear that different climate regions can have varied impacts on crop growth and yield. To account for these regional differences, we divided China into seven sub-regions based on climate conditions and geolocation (see Fig. 1) (Li et al., 2021a; Zhao, 1983). These sub-regions include the temperate and warm-temperate deserts of Northwestern China (Sub-region I), Inner Mongolia (Sub-region II), the temperate humid and sub-humid region of Northeastern China (Sub-region III), the warm-temperate humid and sub-humid region of Northern China (Sub-region IV), subtropical humid Central and Southern China (Sub-region V), the tropical humid region of Southern China (Sub-region VI), and the Qinghai-Tibetan Plateau (Sub-region VII) (Li et al., 2019a). Each of these regions encapsulates distinct climate conditions (Li et al., 2020; Yao et al., 2020), and could influence crop yield differently (Li et al., 2022). This approach allows for a more nuanced

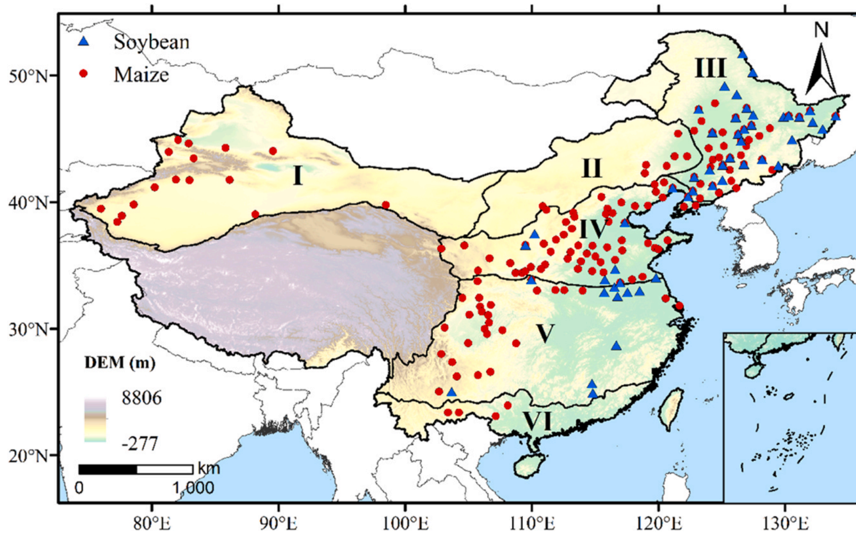

Fig. 1. Location and elevation of field trial sites for maize and soybean yields in China.

understanding of how regional climate conditions impact agricultural productivity. The Qinghai-Tibetan Plateau was excluded from our study as maize and soybean are not primarily grown in this region.

2.2. Data

2.2.1. Crop data

To collect data for our study, we collected maize yield trial data from 155 sites and soybean yield trial data from 50 sites spanning 1999–2010. These data included crop yield and growth stage information (Fig. 1), and was obtained from the China Meteorological Data Sharing Network (http://data.cma.cn/). The management practices and harvest methods used in the studies were consistent with those of local farmers. However, due to gaps in the observed crop yield data from 1999 to 2013, we acquired 1076 sets of field data for maize and 368 sets for soybean.

The risks of CPDs pose a significant threat to food production (Sundström et al., 2014; Trebicki and Finlay, 2019), but CPDs have been largely overlooked in crop yield projections due to modeling and experimental data limitations. Wang et al. (2021a) recently developed a dataset that recorded the occurrence of CPDs in China during historical and future periods. In this study, we developed a CPD occurrence risk model by incorporating a Bayesian approach and utilizing their dataset to account for CPD risk. Detailed information can be found in the supplementary materials and Fig. S1.

2.2.2. Climate data

We obtained climate data from the China Meteorological Data Sharing Network (http://data.cma.cn/), including temperature, precipitation, hours of sunshine, and relative humidity. From these variables, we identified ECEs that could impact crop yield using parameters such as the Standardized Precipitation and Evapotranspiration Index (SPEI), heat (TD, number of days with daily Tmax ≥ 35 °C for maize and Tmax ≥ 30 °C for soybean), cold days (CD, number of days with daily minimum temperature of ≤ 8 °C), heavy rainfall (R30, number of days with precipitation > 30 mm), and drizzle (R1, number of days with precipitation between 0.1 and 1.0 mm). We used threshold values for heat and chilling based on the temperature for the radiation use efficiency for maize and soybean, as outlined by Soltani (2012). We used 21 GCMs that included two climate scenarios (SSP126 and SSP585) to project future ECEs, downloaded from the CMIP6 site (https://esgf-node.llnl.gov/projects/cmip6/).

2.3. Statistical downscaling methods

The NWAI-WG statistical downscaling method is a technique based on a weather generator. To perform spatial analysis, we selected the four grid points nearest to the station and used the inverse distance weighting interpolation method to interpolate values for the station. We used QQ plots to compare observations with simulations and calibrated simulation values using probability distribution functions. Because the climate model data selected for this study were monthly averages, downscaling was necessary to obtain daily data. Generation of precipitation downscaling relied mainly on a gamma function and a first-order Markov chain. We used the first-order Markov chain to generate daily sequences, and the gamma distribution to calculate precipitation probabilities. The density function of the gamma distribution was determined as:

$$f(p) = \frac{p^{\alpha-1} e^{\frac{-p}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)} \qquad p, \quad \beta > 0, \quad 0 < \alpha < 1 \tag{1}$$

We generated temperature data using the WGEN weather generator, and values were calculated similarly to precipitation. The sequence correlation coefficient and cross-correlation coefficient were used:

$$X_i(j)=AX_i-j(j)+\varepsilon_i(j) \tag{2}$$

where X(j) is a matrix containing three climate variables for the *i*th day (including maximum temperature, minimum temperature, and solar radiation), and ε_i is an independent random variable. A and B are matrices defined by the following equation:

$$A = M_1 M_0^{-1}, BB^T = M_0 - AM_1^T \tag{3}$$

where M_0 is the correlation between these three variables on the same day, and M_1 is the correlation with a lag of one day.

Due to the randomness of the probability estimation for precipitation, the sum of daily precipitation data did not always equal the input monthly value. Therefore, we used 1500 iterations to reduce the differences to an acceptable range. This downscaling method is widely used as it is efficient, flexible, and has a good bias correction effect (see Liu and Zuo (2012) for more details).

2.4. GGCM emulators

We projected future crop yields using GGCM emulators developed by Franke et al. (2020b) (GGCM Phase II) that were driven by changing factors from GCMs. The GGCM emulator approach combined the advantages of crop models and statistical models, thereby considering the biological and physical processes of the model and some non-linear responses. The emulators calculated crop yields for each crop and geographic location (at 0.5° resolution) using atmospheric CO_2, changes in growing-season temperature (ΔT) and precipitation (ΔP), and N fertilizer inputs (CTNW-A). The CTNW-A experiment was described in detail in the GGCMI phase 2 description given by Franke et al. (2020a). Several studies have successfully applied this approach to better reproduce global yield changes (Liu et al., 2021a; Zabel et al., 2021). We used GGCM emulators that exhibited high flexibility in capturing geographic differences in crop yield and yield responses and performed well on models with different sensitivities to climate or CO_2 changes. The maize simulations included nine models (CARAIB, EPIC-TAMU, JULES, GEPIC, LPJ-GUESS, LPJmL, pDSSAT, PEPIC, and PROMET), and the soybean simulations included eight models (CARAIB, EPIC-TAMU, JULES, GEPIC, LPJmL, pDSSAT, PEPIC, and PROMET).

Because GGCMs often differ in simulating benchmark crop productivity levels (Müller et al., 2017), we corrected the yield output of crop models to match the observed yield during the observation period, as follows:

$$Y_t^* = Y_{t,c} \times \frac{O_{r,c}}{Y_{r,c}} \tag{4}$$

where Y^*_t is the calibrated yield for time step t, Y_t is the simulated yield for time step *t*, $O_{r,c}$ is the observed yield at station *c* during the historical period *r* (1999–2010 in this study), and $Y_{r,c}$ is the simulated yield during the historical period. This bias correction method helps minimize systematic errors between GGCM and observed data to better align the GGCM predictions with the historical baseline. This approach also improves the accuracy of raw GGCMs and further enhances the relative importance of individual GGCMs within the random forest (RF) model, allowing for better retention of each GGCM's unique characteristics. Thus, we obtained a more robust starting point for the hybrid model (GGCM+RF) and mitigated the influence of inherent biases in the GGCMs.

We used data from 21 monthly GCMs to drive GGCM emulators, and we resampled the GCM data to 0.5° resolution for compatibility. We computed the mean temperature and total precipitation within the growing season by taking the weighted average (sum) of monthly mean temperature data and precipitation data, respectively. The number of days each month during the growing season served as the weight for the calculation.

2.5. Random forest algorithm

We used the RF algorithm to create a hybrid model incorporating ECEs and CPDs with GGCM output. The RF model is well-suited for capturing linear and non-linear relationships between crop yields and environmental factors (Breiman, 2001; Peichl et al., 2021). Moreover, the RF model provides a measure of the relative importance of different predictors, and therefore assists in comprehending how ECEs and CPDs affect crop yields (Feng et al., 2019). Our previous studies have demonstrated the RF model's performance in agriculture (Li et al., 2022; Li et al., 2021a). Overall, incorporating RF with ECEs and CPDs, along with GGCM output, can improve the reliability of crop yield projections.

We used the 'randomForest' package in R software to perform the RF algorithm. We tuned two hyperparameters (n_{tree} and m_{try}) to optimize the RF model by conducting a cross-validation and selecting the smallest RMSE using the 'caret' R package (see Fig. S2). For n_{tree}, we set the range between 100 and 1100, with 200 intervals, and for m_{try}, we set the range between 1 and 7. We ultimately selected $m_{try} = 3$ and $n_{tree} = 1100$ for maize, and $m_{try} = 5$ and $n_{tree} = 900$ for soybean. We used the '%IncMSE' metric to evaluate the relative importance of predictors in the RF model.

2.6. Dynamic linear model

We also used the dynamic linear model (DLM) as part of our hybrid modeling approach to enhance the overall model performance. The DLM is a statistical model widely used in regression analysis, known for its effectiveness in capturing time series characteristics (Prado and West, 2010). The DLM relies on data from the previous time step to estimate states, utilizing observations according to the classical Kalman filter formulas. Recently, this model has been widely used in examining the impacts of climate change on Earth's system and environmental science (Hein et al., 2018; Liu et al., 2019; Zhang et al., 2022; Zhang et al., 2021). Here, we use the DLM as the linear model component of our hybrid models, enabling comparisons with non-linear models (e.g., RF model). The DLM modeling was executed using the 'dynlm' R package (Zeileis et al., 2005).

2.7. Uncertainty analysis

We used analysis of variance (ANOVA) to quantify the primary sources of uncertainty in yield projections for maize and soybean, considering the climate model, crop model, and emission scenario, and using the following formula:

$$SS = SS_{GGCM} + SS_{GCM} + SS_{Scen} + SS_{GCM\times GGCM} + SS_{GCM\times Scen} + SS_{GGCM\times Scen} + SS_{GGCM\times GCM\times Scen} \tag{5}$$

where SS_{GGCM}, SS_{GCM}, and SS_{Scen} are the uncertainty sources from GCM, GGCM, and emission scenarios, and $SS_{GCM\times GGCM}$, $SS_{GCM\times Scen}$, $SS_{GGCM\times Scen}$, and $SS_{GGCM\times GCM\times Scen}$ are their interaction effects.

2.8. Modeling framework

Fig. 2 presents the modeling framework utilized in this study. First, we integrated the outputs from the GGCM-emulator, environmental variables (ECEs and CPDs), and observed crop yield data (collected from 155 sites for maize and 50 sites for soybean) into both the Dynamic Linear Model (DLM) and Random Forest (RF) model. This integration helped in constructing hybrid models. Then, we used the hybrid model to predict crop yields under climate change, driven by future crop yields (computed by GGCM emulators driven by GCM), ECEs (based on GCMs), and CPDs. Finally, we compared crop yield projections and sources of uncertainties between the GGCM and the hybrid models.

2.9. Model performance assessment

We used a leave-one-year-out cross-validation method to assess the model's accuracy in predicting historical crop yields for maize and soybean. In this study, we used Pearson's correlation coefficient (r) and normalized root mean square error (nRMSE) to assess model performance. The equations are written as follows:

$$r = \frac{\sum_{i=1}^{n}(x(i) - \overline{x})(y(i) - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x(i) - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y(i) - \overline{y})^2}} \tag{6}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x(i) - y(i))^2}{n}} \tag{7}$$

$$nRMSE = \frac{RMSE}{\overline{x}} \tag{8}$$

where y(*i*) and *x(i)* are the *i*th simulated and observed yield values (from agricultural meteorological station), respectively; $\overline{y}$ and $\overline{x}$ represents the mean of forecasted and observed values; *n* is the number of samples.

We generated figures using the 'ggplot2' R package (Wickham, 2011) and mapped the spatial distribution of field trial sites using ArcGIS 10.3

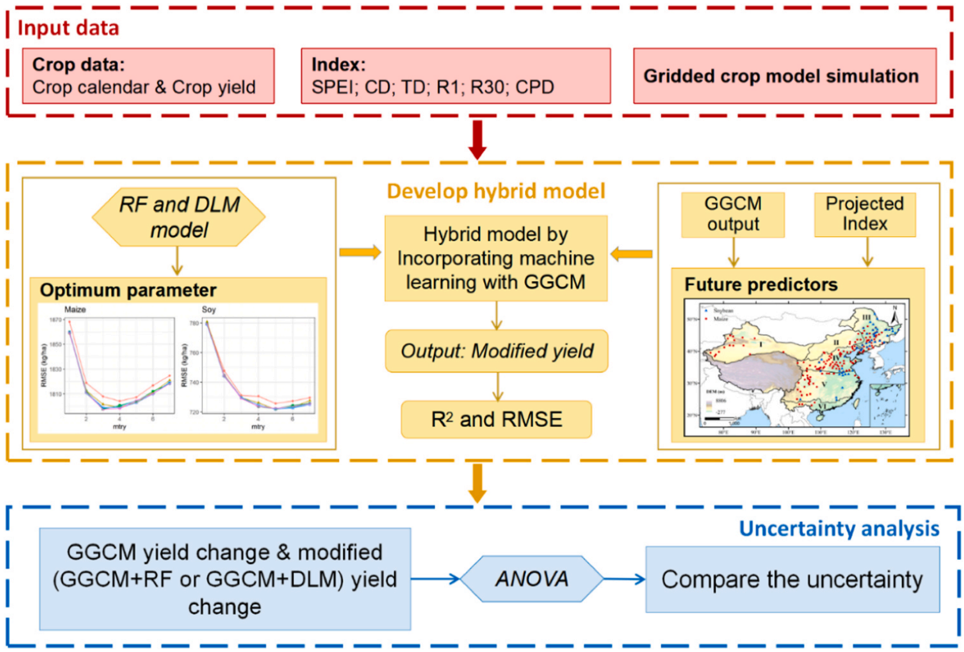

Fig. 2. Framework of the hybrid modeling approach used in this study based on GGCMs and RF model. SPEI, Standardized Precipitation and Evapotranspiration Index; CD, cold days; TD, tropical days; R1, drizzle (daily precipitation between 0.1 and 1 mm); R30, excessive rainfall; CPD, crop pests and diseases. RF, random forest; DLM, dynamic linear model; GGCM, global gridded crop model; R^2, coefficient of determination; RMSE, root mean square error.

software.

3. Results

3.1. Model performance

We assessed the performance of the GGCM and hybrid model (GGCM+RF and GGCM+DLM) using leave-one-year-out cross-validation during the historical period from 1999 to 2010 (Fig. 3). While the GGCMs generally showed responses that were consistent with field experiments (Zhao et al., 2016a), they had poor accuracy at the site scale. Our results indicated that the GGCM had relatively low accuracy in simulating crop yield, with r values ranging from 0.15 to 0.61 and nRMSE ranging from 0.18 to 0.50 for maize, and r values ranging from 0.37 to 0.70 and nRMSE ranging from 0.17 to 0.35 for soybean. However, the hybrid model (GGCM+RF) showed significant improvement, with r ranging from 0.64 to 0.77 and nRMSE ranging from 0.13 to 0.17 for maize, and r ranging from 0.54 to 0.70 and nRMSE ranging from 0.17

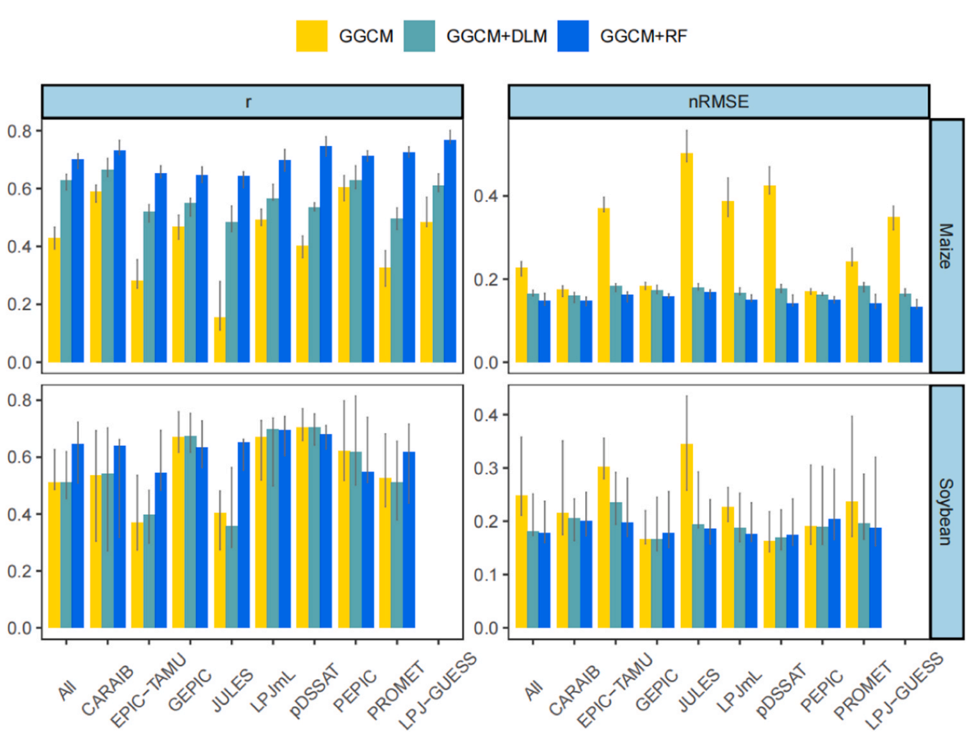

Fig. 3. Model performance of different models and all model ensembles for maize and soybean, based on the evaluation for each (left out) year during 1999–2010. The filled bars represent the mean values of r and nRMSE; the error bars represent the standard errors for the 12 years. GGCM, the model performance of use of GGCM alone; GGCM+DLM, the model performance of hybrid models that integrates GGCM and DLM; GGCM+RF, the model performance of hybrid models that integrates GGCM and RF.

to 0.20 for soybean. As for GGCM+DLM, while performing better than the standalone GGCM, was not as effective as the GGCM+RF model. For maize, its r ranged from 0.48 to 0.67 and nRMSE from 0.17 to 0.19, while for soybean, the r values between 0.36 and 0.70 and nRMSE from 0.17 to 0.2. Thus, the GGCM+RF within the different environmental variables (ECEs and CPDs) can significantly improve the model performance. While previous research has shown that GGCMs capture temperature and drought responses well (Franke et al., 2020a), they may

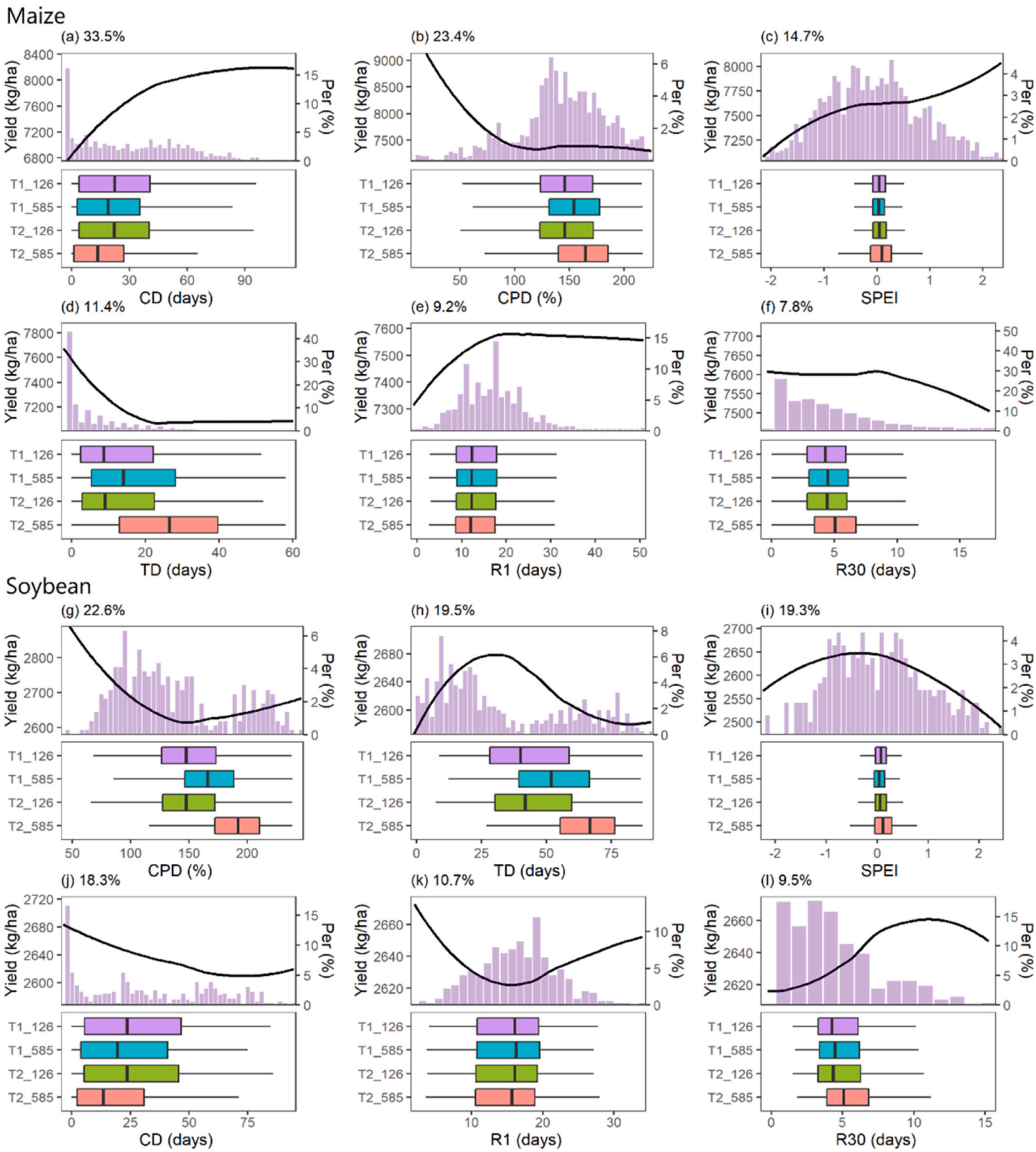

Fig. 4. Partial dependence plots for maize and soybean yields based on six yield predictors. The smoothed black lines represent the model response, with corresponding fitted values for the calibration data. The histograms show the probability distributions for the indexes, while the box plots illustrate the projected frequency of extreme climate events and crop pests and diseases in two future periods (T1: 2040–2069 and T2: 2070–2099) based on 21 downscaled GCMs under SSP126 and SSP585. The percentages shown next to the panel letters indicate the relative importance of each variable generated from the random forest algorithm. SPEI, Standardized Precipitation and Evapotranspiration Index; CD, cold days; TD, tropical days; R1, drizzle (daily precipitation range of 0.1–1 mm); R30, excessive rainfall; CPD, crop pests and diseases.

underestimate the magnitude of yield losses under extreme conditions (Heinicke et al., 2022), and can be insensitive to extreme precipitation events, often overestimating crop yields under extremely wet conditions (Li et al., 2019b). By incorporating such extreme events, our hybrid model could provide more accurate and responsive simulations.

Our analysis demonstrated that certain raw GGCM models performed better than others in predicting crop yields. For maize, pDSSAT and CARAIB produced the best performance among the raw GGCM models, while LPJ-GUESS, PEPIC, pDSSAT, and the multi-model ensemble (designated as “All” in Fig. 3) performed best for the hybrid models (Fig. 3). For soybean, GEPIC and PEPIC produced the best performance among the raw models, while the three EPIC-based models (EPIC-TAMU, GEPIC, PEPIC) and the multi-model ensemble performed best for the hybrid models. However, JULES had poor accuracy for maize and soybean. Our study also revealed that the multi-model average often failed to achieve optimal accuracy, particularly for GGCMs, likely due to the significant differences among models masking some model characteristics and failing to capture all relevant information. As a result, directly averaging the models, such as using an arithmetic average, may not produce satisfactory results when evaluating yield using a multi-model approach. Therefore, selecting appropriate models is crucial for achieving optimal simulation or prediction when considering multiple models.

3.2. Relationship between crop yields and environmental factors

We used the RF model to generate the relative importance and partial dependence plots (PDP) (Fig. 4) to demonstrate the non-linear relationship between environmental factors and crop yield. We normalized the relative importance of the six environmental factors to sum to 100%, and found that the main influencing factors were CD (33.5%) and CPD (23.4%) for maize, and CPD (22.6%) and TD (19.5%) for soybean.

The PDP revealed that cold days (CD) had a positive effect on maize yield but a negative impact on soybean yield, while crop pests and diseases (CPD) significantly reduced yields for both crops. Maize yield demonstrated a positive correlation with the Standardized Precipitation Evapotranspiration Index (SPEI), indicating that the dry season negatively affected yield, whereas the wet season had a positive influence. Soybean yield showed a positive correlation with SPEI ≤ -0.2 and a negative correlation with SPEI > -0.2. Tropical days (TD) negatively impacted both maize and soybean yields. However, when TD was less than 30 days, soybean yield increased. R1 $\leq$ 20 days had a positive effect on maize yield, but R1 $>$ 20 days slightly reduced maize yield. In contrast, R1 negatively influenced soybean yield when it was less than or equal to 16 days, but when R1 exceeded 16 days, soybean yield was positively correlated with R1. R30 exhibited a negative correlation with both maize and soybean yields, but when R30 was below approximately nine days, it was positively correlated with soybean yield (Fig. 4).

Under SSP585, the risk of CD was relatively low compared with the historical period, but the risk associated with CD remained high under SSP126 (Fig. 4). CPD increased significantly in both scenarios, with a greater increase under SSP585 than under SSP126, particularly from 2070 to 2999 (T2). TD had little impact on crop yield under SSP126, but caused significant yield losses under SSP585. R1 did not significantly change under SSP126 or SSP585 compared with the historical data, while R30 increased significantly during T2 under SSP585 (Fig. 4).

Overall, our findings suggested that managing drought and heat events during the growing season will be crucial when cultivating maize and soybean in the context of future climate change. Moreover, our study highlights the increased risk of CPD in the future, emphasizing the need for implementing appropriate management practices (e.g., crop rotation and integrated pest management) to mitigate the negative impact of CPD on crop yield (Lengai et al., 2020; Tariq et al., 2019).

3.3. Projected yield change

Our analysis of the time series of maize and soybean yield changes under SSP126 and SSP585 from 1980 to 2099 (Fig. 5) showed that the GGCMs and hybrid models (GGCM+RF) projected a substantial decline in maize yield, particularly after 2040 (Fig. 5a). Under the SSP126 scenario, the GGCM model projected an increasing trend in soybean yield, while the GGCM+RF model projected a declining trend in soybean yield after 2040. In contrast, under the SSP585 scenario, the GGCM projected a reduction in soybean yield starting around 2050, whereas the GGCM+RF model projected an earlier yield decrease (Fig. 5d).

The GGCM and GGCM+RF simulations projected a decrease in maize yield, with the similarity between the two projections suggesting that the GGCM emulators adequately reflected yield losses under certain extreme climate conditions. However, for soybean, the GGCM may have somewhat overestimated yields due to insufficient consideration of ECEs, while machine learning could capture such impacts. Moreover, the sensitivity of soybean and maize to different climate factors varies, e.g., soybean yields are more sensitive to maximum and minimum temperatures than maize yields (Hoffman et al., 2020), and some process models may oversimplify these responses and thus underestimate the adverse effects of climate change on soybean yield. In contrast, the hybrid models under SSP585 generally produced slightly lower yield changes than GGCMs, and this result could be attributed to the machine learning model capturing the adaptation and resilience strategies employed by farmers in response to extreme events. Such strategies could potentially mitigate the impacts of extreme events, resulting in higher yields than those predicted exclusively by the GGCM (Siebert et al., 2017; Wang et al., 2021b).

The projected crop yield changes varied across different sub-regions. Sub-regions I and IV exhibited more pronounced declines in maize yield, while sub-regions IV and V demonstrated greater yield declines for soybean (Fig. S3). The GGCM+RF simulations estimated slightly smaller yield declines compared with GGCM, particularly in region V for maize yield. Under the SSP126 scenario, GGCM+RF projected marginally larger yield reductions than GGCM. Under the SSP585 scenario, the GGCM simulations in regions IV and V had more pronounced declines in soybean yield than GGCM+RF. However, in sub-region III, GGCM+RF under SSP126 and SSP585 estimated more significant yield decreases than the GGCM simulations (Fig. S4).

3.4. Uncertainties in the two modeling approaches

The GGCM+RF method substantially reduced the uncertainty range of crop yield changes (compared with the raw GGCM simulations) by 32.8–77.8% for maize and by 56.0–67.6% for soybean under the SSP126 and SSP585 scenarios (Fig. 5b, d). Our results for the raw GGCM simulations indicated that the GGCM was the dominant source of uncertainty, while GGCM+RF reduced GGCM-induced uncertainty, resulting in GCM and climate scenario as the main sources of uncertainty for GGCM+RF during T1 and T2, respectively.

The sources of uncertainty in maize and soybean yield projections varied across regions, highlighting their region-specific nature (Figs. S5 and S6). For maize, the main sources of uncertainty were GGCM in sub-regions I and IV. The main sources of uncertainty for maize in sub-regions III and V were GCM during T1 and climate scenario during T2. For soybean, the main sources of uncertainty were crop model in sub-region III, and GCM and climate scenario in sub-regions IV and V.

4. Discussion

4.1. Model performance

We improved the performance of crop models at site scales by incorporating a machine learning algorithm (Fig. 3). The increased improvement was attributed mainly to the external statistical model (RF

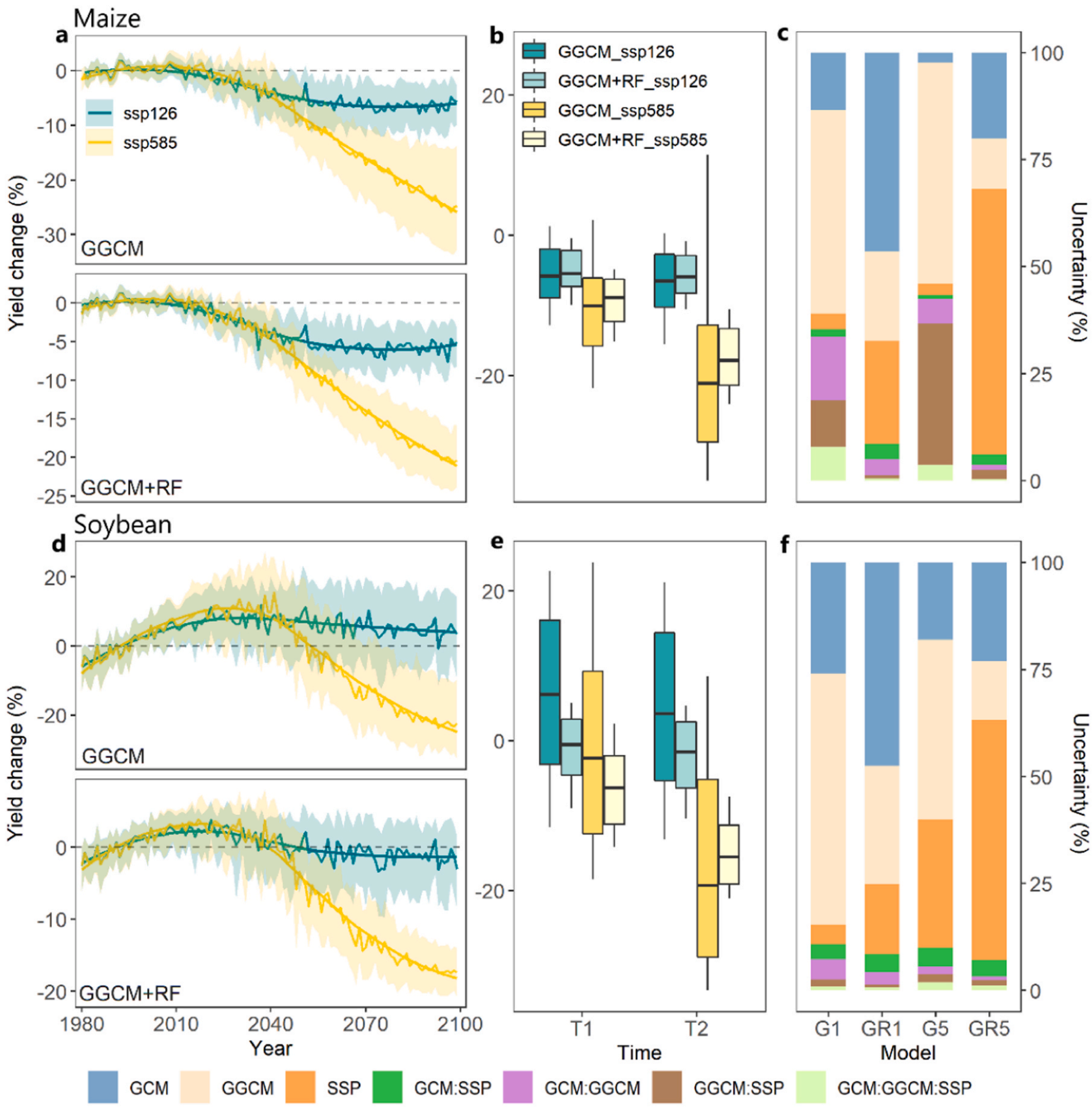

Fig. 5. Projected crop yield changes using multi-model ensembles and the corresponding proportion of uncertainty. **a**, Time series for maize and soybean yield changes for 1980–2099 (left). Shaded areas illustrate the 25th and 75th percentiles. The two solid lines display the median values and a smoothed trend. **b**, Boxplots showing maize and soybean yield changes for 2040–2069 (T1) and 2070–2099 (T2). Box boundaries indicate the 25th and 75th percentiles across 21 GCMs and nine GGCMs; whiskers below and above the box indicate the 10th and 90th percentiles, respectively. **c**, Bar charts showing the source of uncertainty in maize and soybean yield change projections, separated into GCM, GGCM, scenario (SSP), and their interactions. G1, GGCM under SSP126; GR1, GGCM+RF under SSP126; G5, GGCM under SSP585; GR5, GGCM+RF under SSP585.

model) that accounted for the impact of ECEs and CPDs that crop models often oversimplify. When comparing the GGCM+RF and GGCM+DLM models, our results showed superior performance from the GGCM+RF model. Such results are not surprising since non-linear models are generally better at capturing the potential impacts of ECEs and CPDs on crop yields. This can be attributed to the inherent non-linear relationships between crop yield and these variables in the natural environment. However, the accuracy of simulations during the historical period was relatively lower than those using multi-source environmental variables (Cao et al., 2021a; Cao et al., 2021b; Li et al., 2022) because we only used six variables during the growing season and crop model output as predictive factors, and did not consider individual growth stages or additional remote sensing data. We opted for this approach to avoid compromising the distinct features of the GGCM within the machine learning process. Including too many variables may lead to convergence in the results of different models, ultimately weakening their distinct features. Conversely, using too few variables may result in significant model biases. Therefore, finding the optimal number of variables is crucial for improving model accuracy.

4.2. Crop yield responses to environmental factors

We used partial dependence plots to examine the relationships between environmental factors, including ECEs and crop pests and diseases (CPDs), and crop yields (Fig. 4). CPD risk had a significant negative correlation with crop yield. However, maize and soybean exhibited similar yield loss plateaus at CPD risk values > 220% that we attributed to unknown factors that affect crop yields above this threshold. Furthermore, maize and soybean yields exhibited different CD response curves, with CD positively impacting maize, but negatively affecting soybean yield. This result might be due to cool conditions during summer that can reduce heat stress in maize (Bheemanahalli et al., 2022), improve water use efficiency (Zhang et al., 2008), and optimize photosynthesis (Huang et al., 2019). For soybean, CD could slow the rate of photosynthesis or cause frost damage if CD duration is excessively long (Kumar et al., 2018). Notably, the frequency of CD risk is expected to decrease under future climate change, particularly under the SSP585 scenario (Fig. 4), in which heat also leads to large yield losses, especially for maize. Heat impacts maize yield by reducing pollen production (Hatfield and Prueger, 2015) and diminishing leaf area, subsequently affecting photosynthesis (Bheemanahalli et al., 2022).

Our analysis revealed that SPEI was positively correlated with maize yield, indicating that maize yield increases under wetter conditions and decreases under drier conditions. However, for soybean, SPEI was negatively correlated with yield at SPEI > -0.2 (Fig. 4). This type of relationship analysis has been used to examine the water constraint on vegetation productivity (Jiao et al., 2021) Our results demonstrated a negative correlation when SPEI values were relatively higher (SPEI > -0.2), indicating that excessive water could limit soybean growth. This could be attributed to extremely wet conditions causing yield reductions, a response that is often overlooked by crop models (Li et al., 2019b) and explains why crop models generally overestimated soybean yields in our study (Fig. 5). Furthermore, our findings suggested that soybean yields will decline under the R1 scenario, potentially due to drizzle not supplying enough available water for crops while inducing cold and radiation stress (Lesk et al., 2020). Excess rainfall has a more pronounced effect on maize than soybean, possibly because soybean is more sensitive to flooding during the seedling stage (around April–May), while heavy rain is more common in July and August. Heavy rain is often accompanied by strong winds. Under such conditions or other severe convective weather events, maize is more likely to experience lodging or pest and disease infestations (Chaloner et al., 2021; Li et al., 2019b; Savary et al., 2019). Analyzing the relationships between different factors and crop yields helps us better understand the impact of ECEs and CPD risks on crop yields. In general, our study integrated these responses through the use of a hybrid model, successfully improving the model's performance and reducing overall uncertainty.

4.3. Constraining model uncertainty regarding future crop yield changes

Due to the significant uncertainties in current crop models, some studies have attempted to constrain model uncertainty using linear emergent approaches (Wang et al., 2020c; Zhao et al., 2016b), but they often overlook potential non-linear responses. While recent studies have used non-linear models, such as machine learning, to constrain uncertainty (Yin and Leng, 2022), they focused on country-scale analysis rather than field-scale assessments. In addition to non-linear responses, we considered ECEs and CPD risk, and further analyzed sources of uncertainty in raw GGCM and GGCM+RF data to provide a deeper understanding of the impact of future climate change on agriculture.

Understanding the main sources of uncertainty in climate-crop modeling, such as crop models, GCM, and scenarios is essential for formulating effective adaptation strategies for cropping systems. In addition, this understanding can direct future research toward enhancing the precision and effectiveness of yield change predictions (Müller et al., 2021; Wang et al., 2020a). Our findings indicated that GGCM was the main source of uncertainty for raw GGCM simulations, while for GGCM+RF, the dominant source of uncertainty was GCM during T1, while the dominant source was SSP during T2 (Fig. 5). This may be due to our selection of the low and high ends of the Shared Socioeconomic Pathways (SSP126 and SSP585), potentially contributing to a greater variance source from SSP during T2. Furthermore, these sources of uncertainty varied across different maize and soybean regions (Figs. S4 and S5), making accurate crop growth modeling challenging (Jiang et al., 2022; Wang et al., 2020a) and highlighting the need for region-specific modeling approaches. For instance, GGCM was the main source of uncertainty in maize yield changes in sub-regions I and IV, possibly attributed to their semi-arid climate with limited precipitation and significant seasonal temperature variations (Fig. 1). In contrast, GCM was the main source of uncertainty in maize yield changes in sub-regions III and V, characterized by sub-humid continental and subtropical humid climates with relatively high humidity and rainfall (Li et al., 2020; Wang et al., 2022b). As a result, GCMs may struggle to capture distinct climatic patterns and their impacts on agriculture.

4.4. Potential adaptations to mitigate climate change impacts

Consistent with the results of other studies, we found that climate change and global warming will reduce maize and soybean yields (Fig. 5) (Huang et al., 2021; Kothari et al., 2022), particularly after the mid-21st century. Without appropriate adaptation measures, climate change will likely decrease agricultural productivity and increase the risk of crop failure (Huang et al., 2021; Kothari et al., 2022). Developing new crop varieties adapted to differing growing season conditions can help stabilize yields under future climate change (Zabel et al., 2021), especially after the 2040 s (Fig. S7). Adjusting planting dates (Huang et al., 2020; McDonald et al., 2022; Minoli et al., 2022), switching crop species or varieties (Xie et al., 2023), and implementing improved management practices (Liu et al., 2021b; McCullough et al., 2022; Wang et al., 2022d) will be potential strategies for adapting to climate change. Moreover, our findings suggest that optimizing management practices can increase yields while reducing greenhouse gas emissions and achieving climate change adaptation and mitigation goals (He et al., 2022; Liu et al., 2021b; Peng and Guan, 2021).

4.5. Limitations and future framework

Even though the results of our study demonstrated that the uncertainty of multi-model predictions was significantly reduced by incorporating external statistical models, some limitations remain regarding the results of the study. The first potential limitation is the impact of data quality and quantity on the performance of machine learning models (Elavarasan et al., 2018; Liakos et al., 2018). Due to data collection limitations, the observational data used in this study were relatively scarce, and biases in data quality may have affected model prediction results. For instance, during the T2 period under SSP585, some extreme indices exceeded the training dataset range. Predictions made beyond this range could be unreliable, resulting in slight overestimations of crop yields simulated by GGCM+RF (Fig. 5). Moreover, the coarse data resolution for CPDs (provincial scale) may influence the prediction accuracy. Thus, more controlled variable experiments are needed to better capture the impact of ECEs on crop yields. The second potential limitation relative to the results of our study arises from the fact that our research focused on using external statistical models to improve model prediction accuracy and to constrain uncertainty rather than enhancing the response processes used in the crop models. Other studies have attempted to improve crop models to enhance model performance and reduce uncertainty by addressing the temperature response (Wang et al., 2017; Wang et al., 2020b), the photosynthetic response (Wang et al., 2022c), and the waterlogging response (Liu et al., 2023). The third limitation is due to the fact that considerable differences in climate change response processes among crop models could significantly influence the sources of uncertainty and yield projections. Future studies should consider model composition and ensemble size when comprehensively analyzing crop yield changes and associated uncertainty. Lastly, it is important to differentiate between uncertainty and inaccuracy. While our study successfully reduced the overall uncertainty of crop yield projections, further efforts are needed to enhance the accuracy of crop models. These effort will call for additional experiments to better harmonize crop and climate model processes, ultimately increasing the overall robustness and providing more reliable information for decision-making and future research.

5. Conclusion

In summary, we successfully developed hybrid models (GGCM+RF) that combined machine learning with crop models by considering ECEs and CPD, thereby improving the accuracy of maize and soybean yield projections while reducing the overall uncertainty. Our main conclusions are:

1) Incorporating machine learning into crop models significantly improved model performance (higher r values and lower nRMSE values), particularly for maize, with r values increasing from 0.15 to 0.61–0.64–0.77, and nRMSE values decreasing from 0.18 to 0.50–0.13–0.17.
2) The dominant factors affecting yield changes in China were CD, CPD, and SPEI for maize, and CPD, tropical days (TD), and SPEI for soybean. In addition, our findings highlighted the impact of ECEs and CPDs on crop yields, demonstrating that crop yields exhibit threshold-like responses to such environmental variables.
3) Our approach of combining machine learning with crop models reduced yield uncertainty by 32.8–77.8% for maize and 56.0–67.6% for soybean. For the raw GGCM, the crop model was the main source of uncertainty, while for the GGCM+RF model, the main sources of uncertainty were GCM during T1 and SSP during T2.

The results of our study provide valuable information for local farmers and policymakers to address the impacts of more extreme climate under future climate change. We anticipate that our methods can be extended to other regions or applied globally.

CRediT authorship contribution statement

Linchao Li: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Yan Zhang**: Conceptualization, Data curation. **Bin Wang:** Formal analysis, Validation. **Puyu Feng:** Formal analysis, Visualization. **Qinsi He**: Investigation, Methodology. **Yu Shi:** Methodology, Software. **Ke Liu:** Investigation, Software. **Matthew Tom Harrison**: Investigation, Writing – original draft, Writing – review & editing. **De Li Liu:** Conceptualization, Funding acquisition, Resources, Supervision, Writing – original draft. **Ning Yao**: Visualization, Writing – original draft, Writing – review & editing. **Yi Li**: Project administration, Writing – review & editing. **Jianqiang He:** Formal analysis, Writing – review & editing. **Hao Feng**: Resources, Writing – review & editing. **Kadambot H.M. Siddique:** Validation, Writing – original draft, Writing – review & editing. **Qiang Yu**: Conceptualization, Funding acquisition, Project administration, Resources, Supervision.

Declaration of Competing Interest

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

Acknowledgments

This study was supported by the Natural Science Foundation of China (no. 41961124006). We acknowledge the Agricultural Model Intercomparison and Improvement Project (AgMIP) and Inter-Sectoral Impact Model Intercomparison Project (ISIMIP). We thank the GGCMI research team for producing and making their GGCM CTWN-A simulations available. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modeling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making their model output available, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies that support CMIP6 and ESGF. Additionally, we appreciate the support provided by the Fundamental Research Funds for the 111 Project [B12007]. The NSW Department of Primary Industries provided office facilities.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.eja.2023.126917.

References

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., et al., 2013. Uncertainty in simulating wheat yields under climate change. Nat. Clim. Change 3 (9), 827–832.
Asseng, S., Ewert, F., Martre, P., Rötter, R.P., Lobell, D.B., et al., 2014. Rising temperatures reduce global wheat production. Nat. Clim. Change 5 (2), 143–147.
Balkovič, J., van der Velde, M., Schmid, E., Skalský, R., Khabarov, N., et al., 2013. Pan-European crop modelling with EPIC: Implementation, up-scaling and regional crop yield validation. Agric. Syst. 120, 61–75.
Bheemanahalli, R., Ramamoorthy, P., Poudel, S., Samiappan, S., Wijewardane, N., et al., 2022. Effects of drought and heat stresses during reproductive stage on pollen germination, yield, and leaf reflectance properties in maize (Zea mays L.). Plant Direct 6 (8), e434.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Cao, J., Zhang, Z., Luo, Y., Zhang, L., Zhang, J., et al., 2021a. Wheat yield predictions at a county and field scale with deep learning, machine learning, and google earth engine. Eur. J. Agron. 123.
Cao, J., Zhang, Z., Tao, F., Zhang, L., Luo, Y., et al., 2021b. Integrating multi-source data for rice yield prediction across China using machine learning and deep learning approaches. Agric. For. Meteorol. 297.
Chaloner, T.M., Gurr, S.J., Bebber, D.P., 2021. Plant pathogen infection risk tracks global crop yields under climate change. Nat. Clim. Change 11 (8), 710–715.
Deutsch, C.A., Tewksbury, J.J., Tigchelaar, M., Battisti, D.S., Merrill, S.C., et al., 2018. Increase in crop losses to insect pests in a warming climate. Science 361 (6405), 916–919.
Elavarasan, D., Vincent, D.R., Sharma, V., Zomaya, A.Y., Srinivasan, K., 2018. Forecasting yield by integrating agrarian factors and machine learning models: a survey. Comput. Electron. Agric. 155, 257–282.
FAO, 2020. FAOSTAT database. Food and Agriculture Organization of the United Nations. (⟨http://www.fao.org/faostat/en/#data/QC⟩).
Feng, P., Wang, B., Liu, D.L., Waters, C., Yu, Q., 2019. Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia. Agric. For. Meteorol. 275, 100–113.
Feng, P., Wang, B., Liu, D.L., Waters, C., Xiao, D., et al., 2020. Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique. Agric. For. Meteorol. 285–286.
Franke, J.A., Müller, C., Elliott, J., Ruane, A.C., Jägermeyr, J., et al., 2020a. The GGCMI Phase 2 experiment: global gridded crop model simulations under uniform changes in CO_2, temperature, water, and nitrogen levels (protocol version 1.0). Geoscientific Model. Development 13 (5), 2315–2336.
Franke, J.A., Müller, C., Elliott, J., Ruane, A.C., Jägermeyr, J., et al., 2020b. The GGCMI Phase-2 emulators: global gridded crop model responses to changes in CO_2, temperature, water, and nitrogen (version 1.0). Geosci. Model Dev. 13 (9), 3995–4018.
Harrison, M.T., 2021. Climate change benefits negated by extreme heat. Nat. Food 2 (11), 855–856.
Harrison, M.T., Roggero, P.P., Zavattaro, L., 2019. Simple, efficient and robust techniques for automatic multi-objective function parameterisation: Case studies of local and global optimisation using APSIM. Environ. Model. Softw. 117, 109–133.
Hatfield, J.L., Prueger, J.H., 2015. Temperature extremes: effect on plant growth and development. Weather Clim. Extrem. 10, 4–10.
He, Q., Liu, D.L., Wang, B., Li, L., Cowie, A., et al., 2022. Identifying effective agricultural management practices for climate change adaptation and mitigation: A win-win strategy in South-Eastern Australia. Agric. Syst. 203.
Hein, A.M., Gil, M.A., Twomey, C.R., Couzin, I.D., Levin, S.A., 2018. Conserved behavioral circuits govern high-speed decision-making in wild fish shoals. Proc. Natl. Acad. Sci. USA 115 (48), 12224–12228.
Heinicke, S., Frieler, K., Jägermeyr, J., Mengel, M., 2022. Global gridded crop models underestimate yield responses to droughts and heatwaves. Environ. Res. Lett. 17, 4.
Hoffman, A., R Kemanian, A., E Forest, C., 2020. The response of maize, sorghum, and soybean yield to growing-phase climate revealed with machine learning. Environ. Res. Lett. 15, 9.
Huang, M., Piao, S., Ciais, P., Penuelas, J., Wang, X., et al., 2019. Air temperature optima of vegetation productivity across global biomes. Nat. Ecol. Evol. 3 (5), 772–779.
Huang, M., Wang, J., Wang, B., Liu, D.L., Yu, Q., et al., 2020. Optimizing sowing window and cultivar choice can boost China's maize yield under 1.5 °C and 2 °C global warming. Environ. Res. Lett. 15, 2.
Huang, M., Wang, J., Wang, B., Liu, D.L., Feng, P., et al., 2021. Assessing maize potential to mitigate the adverse effects of future rising temperature and heat stress in China. Agric. For. Meteorol. 311.
Huang, M., Wang, J., Wang, B., Liu, D.L., Feng, P., et al., 2022. Dominant sources of uncertainty in simulating maize adaptation under future climate scenarios in China. Agric. Syst. 199.

Jägermeyr, J., Müller, C., Ruane, A.C., Elliott, J., Balkovic, J., et al., 2021. Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. Nat. Food.
Jiang, H., Hu, H., Zhong, R., Xu, J., Xu, J., et al., 2020. A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: A case study of the US Corn Belt at the county level. Glob. Change Biol. 26 (3), 1754–1766.
Jiang, T., Wang, B., Xu, X., Cao, Y., Liu, D.L., et al., 2022. Identifying sources of uncertainty in wheat production projections with consideration of crop climatic suitability under future climate. Agric. For. Meteorol. 319.
Jiao, W., Wang, L., Smith, W.K., Chang, Q., Wang, H., et al., 2021. Observed increasing water constraint on vegetation growth over the last three decades. Nat. Commun. 12 (1), 3777.
Kothari, K., Battisti, R., Boote, K.J., Archontoulis, S.V., Confalone, A., et al., 2022. Are soybean models ready for climate change food impact assessments? Eur. J. Agron. 135.
Kumar, R., Singh, P., Singh, S., 2018. A review report: low temperature stress for crop production. Int. J. Pure Appl. Biosci. 6 (2), 575–598.
Langworthy, A.D., Rawnsley, R.P., Freeman, M.J., Pembleton, K.G., Corkrey, R., et al., 2018. Potential of summer-active temperate (C3) perennial forages to mitigate the detrimental effects of supraoptimal temperatures on summer home-grown feed production in south-eastern Australian dairying regions. Crop Pasture Sci. 69 (8), 808–820.
Lengai, G.M., Muthomi, J.W., Mbega, E.R., 2020. Phytochemical activity and role of botanical pesticides in pest management for sustainable agricultural crop production. Sci. Afr. 7, e00239.
Lesk, C., Rowhani, P., Ramankutty, N., 2016. Influence of extreme weather disasters on global crop production. Nature 529 (7584), 84–87.
Lesk, C., Coffel, E., Horton, R., 2020. Net benefits to US soy and maize yields from intensifying hourly rainfall. Nat. Clim. Change 10 (9), 819–822.
Li, L., Yao, N., Liu, D.L., Song, S., Lin, H., et al., 2019a. Historical and future projected frequency of extreme precipitation indicators using the optimized cumulative distribution functions in China. J. Hydrol. 579.
Li, L., Zou, Y., Li, Y., Lin, H., Liu, D.L., et al., 2020. Trends, change points and spatial variability in extreme precipitation events from 1961 to 2017 in China. Hydrol. Res.
Li, L., Wang, B., Feng, P., Wang, H., He, Q., et al., 2021a. Crop yield forecasting and associated optimum lead time analysis based on multi-source environmental data across China. Agric. For. Meteorol. 308–309.
Li, L., Wang, B., Feng, P., Li Liu, D., He, Q., et al., 2022. Developing machine learning models with multi-source environmental data to predict wheat yield in China. Comput. Electron. Agric. 194.
Li, Y., Guan, K., Schnitkey, G.D., DeLucia, E., Peng, B., 2019b. 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.
Li, Z., Zhang, Z., Zhang, L., 2021b. Improving regional wheat drought risk assessment for insurance application by integrating scenario-driven crop model, machine learning, and satellite data. Agric. Syst. 191.
Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D., 2018. Machine learning in agriculture: a review. Sens. (Basel) 18, 8.
Liu, B., Asseng, S., Müller, C., Ewert, F., Elliott, J., et al., 2016. Similar estimates of temperature impacts on global wheat yield by three independent methods. Nat. Clim. Change 6 (12), 1130–1136.
Liu, D.L., Zuo, H., 2012. Statistical downscaling of daily climate variables for climate change impact assessment over New South Wales, Australia. Clim. Change 115 (3–4), 629–666.
Liu, K., Harrison, M.T., Shabala, S., Meinke, H., Ahmed, I., et al., 2020. The state of the art in modeling waterlogging impacts on plants: what do we know and what do we need to know. Earth's. Future 8 (12) e2020EF001801.
Liu, K., Harrison, M.T., Yan, H., Liu, D.L., Meinke, H., et al., 2023. Silver lining to a climate crisis in multiple prospects for alleviating crop waterlogging under future climates. Nat. Commun. 14, 1.
Liu, W., Ye, T., Jägermeyr, J., Müller, C., Chen, S., et al., 2021a. Future climate change significantly alters interannual wheat yield variability over half of harvested areas. Environ. Res. Lett. 16, 9.
Liu, Y., Kumar, M., Katul, G.G., Porporato, A., 2019. Reduced resilience as an early warning signal of forest mortality. Nat. Clim. Change 9 (11), 880–885.
Liu, Z., Ying, H., Chen, M., Bai, J., Xue, Y., et al., 2021b. Optimization of China's maize and soy production can ensure feed sufficiency at lower nitrogen and carbon footprints. Nat. Food 2 (6), 426–433.
Lobell, D.B., Bänziger, M., Magorokosho, C., Vivek, B., 2011. Nonlinear heat effects on African maize as evidenced by historical yield trials. Nat. Clim. Change 1 (1), 42–45.
Lobell, D.B., Roberts, M.J., Schlenker, W., Braun, N., Little, B.B., et al., 2014. Greater sensitivity to drought accompanies maize yield increase in the U.S. Midwest. Science 344 (6183), 516–519.
Mbow, C., Rosenzweig, C., Barioni, L.G., Benton, T.G., Herrero, M. et al., 2019. Food security, Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security and greenhouse gas fluxes in terrestrial ecosystems. IPCC.
McCullough, E.B., Quinn, J.D., Simons, A.M., 2022. Profitability of climate-smart soil fertility investment varies widely across sub-Saharan Africa. Nature. Food 3 (4), 275–285.
McDonald, A.J., Balwinder, S., Keil, A., Srivastava, A., Craufurd, P., et al., 2022. Time management governs climate resilience and productivity in the coupled rice–wheat cropping systems of eastern India. Nat. Food 3 (7), 542–551.
Minoli, S., Jagermeyr, J., Asseng, S., Urfels, A., Muller, C., 2022. Global crop yields can be lifted by timely adaptation of growing periods to climate change. Nat. Commun. 13 (1), 7079.
Müller, C., Elliott, J., Chryssanthacopoulos, J., Arneth, A., Balkovic, J., et al., 2017. Global gridded crop model evaluation: benchmarking, skills, deficiencies and implications. Geosci. Model Dev. 10 (4), 1403–1422.
Müller, C., Franke, J., Jägermeyr, J., Ruane, A.C., Elliott, J., et al., 2021. Exploring uncertainties in global crop yield projections in a large ensemble of crop models and CMIP5 and CMIP6 climate scenarios. Environ. Res. Lett. 16, 3.
Osborne, T., Gornall, J., Hooker, J., Williams, K., Wiltshire, A., et al., 2015. JULES-crop: a parametrisation of crops in the joint UK land environment simulator. Geosci. Model Dev. 8 (4), 1139–1155.
Peichl, M., Thober, S., Samaniego, L., Hansjürgens, B., Marx, A., 2021. Machine-learning methods to assess the effects of a non-linear damage spectrum taking into account soil moisture on winter wheat yields in Germany. Hydrol. Earth Syst. Sci. 25 (12), 6523–6545.
Peng, B., Guan, K., 2021. Harmonizing climate-smart and sustainable agriculture. Nat. Food 2 (11), 853–854.
Piao, S., Ciais, P., Huang, Y., Shen, Z., Peng, S., et al., 2010. The impacts of climate change on water resources and agriculture in China. Nature 467 (7311), 43–51.
Prado, R. and West, M., 2010. Time series: modeling, computation, and inference. Chapman and Hall/CRC.
Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Muller, C., et al., 2014. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. Proc. Natl. Acad. Sci. USA 111 (9), 3268–3273.
Savary, S., Willocquet, L., Pethybridge, S.J., Esker, P., McRoberts, N., et al., 2019. The global burden of pathogens and pests on major food crops. Nat. Ecol. Evol. 3 (3), 430–439.
Siebert, S., Webber, H., Zhao, G., Ewert, F., 2017. Heat stress is overestimated in climate impact studies for irrigated agriculture. Environ. Res. Lett. 12, 5.
Soltani, A., 2012. Modeling Physiology of Crop Development, Growth and Yield. CABi.
Sundström, J.F., Albihn, A., Boqvist, S., Ljungvall, K., Marstorp, H., et al., 2014. Future threats to agricultural food production posed by environmental degradation, climate change, and animal and plant diseases–a risk analysis in three economic and climate settings. Food Secur. 6, 201–215.
Tao, F., Yokozawa, M., Liu, J., Zhang, Z., 2008. Climate–crop yield relationships at provincial scales in China and the impacts of recent climate trends. Clim. Res. 38, 83–94.
Tao, F., Zhang, Z., Xiao, D., Zhang, S., Rötter, R.P., et al., 2014. Responses of wheat growth and yield to climate change in different climate zones of China, 1981–2009. Agric. For. Meteorol. 189–190, 91–104.
Tao, R., Zhao, P., Wu, J., Martin, N.F., Harrison, M.T. et al., 2022. Optimizing Crop Management with Reinforcement Learning and Imitation Learning.
Tariq, M., Ali, H., Hussain, N., Nasim, W., Mubeen, M., et al., 2019. Fundamentals of crop rotation in agronomic management. Agron. Crops: Prod. Technol. Volume 1, 545–559.
Trebicki, P., Finlay, K., 2019. Pests and Diseases under Climate Change; Its Threat to Food Security. John Wiley & Sons Ltd, Chichester.
Wang, B., Feng, P., Liu, D.L., O'Leary, G.J., Macadam, I., et al., 2020a. Sources of uncertainty for wheat yield projections under future climate are site-specific. Nat. Food 1 (11), 720–728.
Wang, C., Wang, X., Jin, Z., Müller, C., Pugh, T.A.M., et al., 2021a. Occurrence of crop pests and diseases has largely increased in China since 1970. Nat. Food 3 (1), 57–65.
Wang, E., Martre, P., Zhao, Z., Ewert, F., Maiorano, A., et al., 2017. The uncertainty of crop yield projections is reduced by improved temperature response functions. Nat. Plants 3, 17102.
Wang, L., Jiao, W., MacBean, N., Rulli, M.C., Manzoni, S., et al., 2022a. Dryland productivity under a changing climate. Nat. Clim. Change 12 (11), 981–994.
Wang, L., Li, Y., Li, M., Li, L., Liu, F., et al., 2022b. Projection of precipitation extremes in China's mainland based on the statistical downscaled data from 27 GCMs in CMIP6. Atmos. Res. 280.
Wang, X., Wang, S., Li, X., Chen, B., Wang, J., et al., 2020b. Modelling rice yield with temperature optima of rice productivity derived from satellite NIRv in tropical monsoon area. Agric. For. Meteorol. 294.
Wang, X., Zhao, C., Müller, C., Wang, C., Ciais, P., et al., 2020c. Emergent constraint on crop yield response to warmer temperature from field experiments. Nat. Sustain. 3 (11), 908–916.
Wang, X., Muller, C., Elliot, J., Mueller, N.D., Ciais, P., et al., 2021b. Global irrigation contribution to wheat and maize yield. Nat. Commun. 12 (1), 1235.
Wang, Y., Liu, Z., Yu, Q., Liu, L., Liu, X., et al., 2022c. Simulations of solar-induced chlorophyll fluorescence over crop canopies using the integrated APSIM model. Comput. Electron. Agric. 203.
Wang, Z., Yin, Y., Wang, Y., Tian, X., Ying, H., et al., 2022d. Integrating crop redistribution and improved management towards meeting China's food demand with lower environmental costs. Nat. Food 3 (12), 1031–1039.
Wickham, H., 2011. ggplot2. Wiley Interdisciplinary Reviews: Computational Statistics, 3(2): 180–185.
Xie, W., Zhu, A., Ali, T., Zhang, Z., Chen, X., et al., 2023. Crop switching can enhance environmental sustainability and farmer incomes in China. Nature.
Xiong, W., Asseng, S., Hoogenboom, G., Hernandez-Ochoa, I., Robertson, R., et al., 2019. Different uncertainty distribution between high and low latitudes in modelling warming impacts on wheat. Nat. Food 1 (1), 63–69.
Yao, N., Li, L., Feng, P., Feng, H., Li, L., et al., 2020. Projections of drought characteristics in China based on a standardized precipitation and evapotranspiration index and multiple GCMs. Sci. Total Environ. 704, 135245.
Yin, X., Leng, G., 2022. Observational constraint of process crop models suggests higher risks for global maize yield under climate change. Environ. Res. Lett. 17, 7.
Zabel, F., Müller, C., Elliott, J., Minoli, S., Jägermeyr, J., et al., 2021. Large potential for crop production adaptation depends on available future varieties. Glob. Change Biol.

Zeileis, A., Leisch, F., Kleiber, C., Hornik, K., 2005. Monitoring structural change in dynamic econometric models. J. Appl. Econ. 20 (1), 99–121.

Zhang, X., Chen, S., Sun, H., Pei, D., Wang, Y., 2008. Dry matter, harvest index, grain yield and water use efficiency as affected by water supply in winter wheat. Irrig. Sci. 27 (1), 1–10.

Zhang, Y., Keenan, T.F., Zhou, S., 2021. Exacerbated drought impacts on global ecosystems due to structural overshoot. Nat. Ecol. Evol. 5 (11), 1490–1498.

Zhang, Y., Gentine, P., Luo, X., Lian, X., Liu, Y., et al., 2022. Increasing sensitivity of dryland vegetation greenness to precipitation due to rising atmospheric CO_2. Nat. Commun. 13 (1), 4875.

Zhao, C., Piao, S., Huang, Y., Wang, X., Ciais, P., et al., 2016a. Field warming experiments shed light on the wheat yield response to temperature in China. Nat. Commun. 7, 13530.

Zhao, C., Piao, S., Wang, X., Huang, Y., Ciais, P., et al., 2016b. Plausible rice yield losses under future climate warming. Nat. Plants 3, 16202.

Zhao, S., 1983. A new scheme for comprehensive physical regionalization in China. Acta Geogr. Sin. 38 (1), 1–10.