**LETTER • OPEN ACCESS**

# Planning maize hybrids adaptation to future climate change by integrating crop modelling with machine learning

View the article online for updates and enhancements.

# ENVIRONMENTAL RESEARCH
## LETTERS

**LETTER**

CrossMark

# Planning maize hybrids adaptation to future climate change by integrating crop modelling with machine learning

Liangliang Zhang[1], Zhao Zhang[1,*], Fulu Tao[2,3,*], Yuchuan Luo[1], Juan Cao[1], Ziyue Li[1], Ruizhi Xie[4] and Shaokun Li[4]

1 Academy of Disaster Reduction and Emergency Management Ministry of Emergency Management and Ministry of Education, School of National Safety and Emergency Management, Beijing Normal University, Beijing 100875, People's Republic of China
2 Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, People's Republic of China
3 College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China
4 Institute of Crop Sciences, Chinese Academy of Agricultural Sciences/Key Laboratory of Crop Physiology and Ecology, Ministry of Agriculture, Beijing 100081, People's Republic of China
* Authors to whom any correspondence should be addressed.

E-mail: zhangzhao@bnu.edu.cn and taofl@igsnrr.ac.cn

## Abstract

Crop hybrid improvement is an efficient and environmental-friendly option to adapt to climate change and increase grain production. However, the adaptability of existing hybrids to a changing climate has not been systematically investigated. Therefore, little is known about the appropriate timing of hybrid adaptation. Here, using a novel hybrid model which coupled CERES-Maize with machine learning, we critically investigated the impacts of climate change on maize productivity with an ensemble of hybrid-specific estimations in China. We determined when and where current hybrids would become unviable and hybrid adaptation need be implemented, as well as which hybrid traits would be desirable. Climate change would have mostly negative impacts on maize productivity, and the magnitudes of yield reductions would highly depend on the growth cycle of the hybrids. Hybrid replacement could partially, but not completely, offset the yield loss caused by projected climate change. Without adaptation, approximately 53% of the cultivation areas would require hybrid renewal before 2050 under the RCP 4.5 and RCP 8.5 emission scenarios. The medium-maturing hybrids with a long grain-filling duration and a high light use efficiency would be promising, although the ideotypic traits could be different for a specific environment. The findings highlight the necessity and urgency of breeding climate resilient hybrids, providing policy-makers and crop breeders with the early signals of when, where and what hybrids will be required, which stimulate proactive investment to facilitate breeding. The proposed crop modelling approach is scalable, largely data-driven and can be used to tackle the longstanding problem of predicting hybrids' future performance to accelerate development of new crop hybrids.

## 1. Introduction

Global food demand is expected to double by 2050, and maize is a staple food for more than 4.5 billion people (Ray *et al* 2012). China is the 2nd largest maize producer and consumer, contributing to 21% of the global production (Faostat 2016). A decline in its productivity could have profound implications for global food security. Climate change may pose a notable threat to food availability and stability by exposing crops to more heat, drought, and other extreme climate events (Lesk *et al* 2016). A recent meta-analysis reported that the national yield of China would decrease by 8% with a warming of 1 °C (Zhao *et al* 2017). Therefore, how crop production will adapt to future climate change has become a key concern (Challinor *et al* 2016, Parent *et al* 2018).

Crop hybrid improvement is an efficient and environmental-friendly option by which agricultural systems can effectively adapt to climate change (Bailey-Serres *et al* 2019). Crop breeders across the globe have long been working on the improvement of crop hybrid, but it remains unknown about when and where the adaptability of current hybrids would be broken and hybrid adaptation should be triggered. Moreover, the traits which minimize yield loss under climate change are normally complex, polygenic and heritability is low, and consequently multiple cycles and decades of selection are necessary. It takes about 20 years or more to develop a new crop hybrid, which involves breeding, delivery and adoption processes and is irreversible and requires substantial investment (Challinor *et al* 2016). Hence, accurate information on the time of hybrid adaptation and the ideotypic traits for a specific environment should be timely identified so as not to spend funds in vain (Morris *et al* 2003, Challinor *et al* 2016, Ramirez-Villegas *et al* 2018), considered the complex interactions between genotype (G), environment (E) and management (M). Investigating critically the responses of existing hybrids to climate change is a basis to identify when and where hybrid adaptation is required, as well as which hybrid traits are promising, for a specific environment.

Many previous studies have investigated climate change impacts and adaptations for crop productivity via field experiments (Zheng *et al* 2018, Yang *et al* 2019), crop model simulations (Li *et al* 2016, Xiao *et al* 2020), and statistical regression (Stevens and Madani 2016). However, these approaches had some limitations. First, the assessments were commonly based on a single crop hybrid rather than several popular hybrids, which might bias the results due to large differences in sensitivity of hybrids to changes in environments and agronomic managements (Shew *et al* 2020). Second, crop models could generally capture key crop growth processes and the G × E × M interactions fairly well at a field scale. However, the high data requirements and computational expense prevent their sound applications over a large area (Tao *et al* 2009, Burke and Lobell 2017, Tao *et al* 2018, Muller *et al* 2019). Moreover, crop models have uncertainties in up-scaling, which may bring bias to results (Martre *et al* 2015, Feng *et al* 2019). Statistical methods link crop yields in given locations to local climate variables and then are used to predict yields under altered climate conditions, which are relatively easy to compute. As an immediate successor of statistical methods, machine learning approach (ML) adopts important weights rather than the likelihood or probability of any information to be forecasted (Crane-Droesch 2018, Reichstein *et al* 2019). ML is capable of disentangling the effects of co-linear climate variables and analysing hierarchical and nonlinear relationships between the predictors and the response variable through an ensemble learning

approach. ML has additional advantages on spatial generalization and lower requirements for computation resources (Grimm *et al* 2008, Folberth *et al* 2019). However, statistical methods obscure the underlying physiological mechanisms on how plants respond to climate change and might sometimes be vague in aiding targeted development of adaptive practices. In complement, developing hybrid models by integrating process-based models with ML models can not only build crop models' capacity in representing the interactions between G × E × M, increase computational efficiency, but also improve robustness in spatial generalization or up-scaling of the simulation results from process-based models (Azzari *et al* 2017, Feng *et al* 2019, Reichstein *et al* 2019).

In this study, we systematically evaluated the impacts of projected climate change on existing maize hybrids and critically identified the timescale and hybrid ideotypic traits for hybrid adaptation in the major maize cultivation regions across China. We first calibrated the cropping system model—CERES-Maize of decision support system for agrotechnology transfer (DSSAT) model based on a large number of field hybrid trials. Then, we simulated maize growth and productivity under massive scenarios of realistic ranges of climate, soil, hybrids, and management practices. The simulated results were applied to train random forest (RF) and eXtreme gradient boosting (XGBoost) models to develop an optimal hybrid model for each typical agro-ecological zone. Next, we drove the optimal hybrid model by a range of projected climate change scenarios for the 2030s and 2050s under the RCP 4.5 and RCP 8.5 emission pathways to investigate the impacts of projected climate change on maize productivity. Finally, we identified the timescale for hybrid adaptation through threshold-crossing analyses, as well as the hybrid ideotypic traits. We aim to answer: (a) do the impacts of climate change on maize productivity depend on crop hybrids? (b) Are the existing hybrids able to adapt to future climate change? (c) If not, when and where will the adaptability be impaired, and furthermore, when should hybrid adaptation be triggered? (d) Which hybrid traits will be promising for adapting to climate change?

## 2. Material and methods

### 2.1. Data

*2.1.1. Field experiments, soil and meteorological data*
A number of maize hybrid trials were conducted by agricultural experts in the Science and Technology Innovation Project of Improving Food Yield and Efficiency. The hybrid trials were carried out at 13 sites in zone I, 13 sites in zone II, and 7 sites in zone III (figure 1 and supplementary text 1 (available online at stacks.iop.org/ERL/16/124043/mmedia)) from 2009 to 2017, which covered 15 provinces and 15 soil types in the Chinese maize belt.
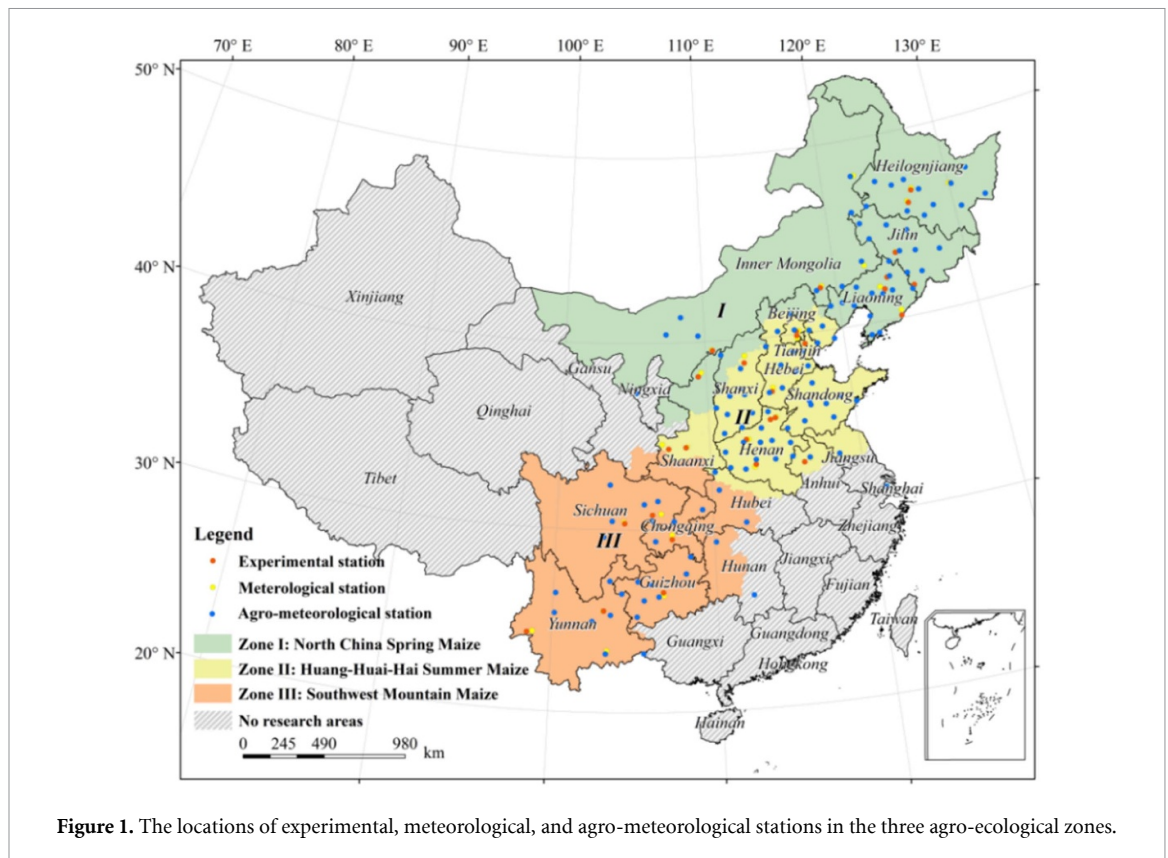
**Figure 1.** The locations of experimental, meteorological, and agro-meteorological stations in the three agro-ecological zones.

Six latest, most commonly used and widely promoted hybrids with three maturity traits were sowed for more than three years in each agro-ecological zone. Various planting densities including 4.5, 6.0, 6.5, and 7.5 plants m$^{-2}$ were experimented across the Chinese maize belt. The sowing dates, row spacing, irrigation and fertilization applications were consistent with the practices of local farmers. The hybrid characteristic (i.e. maturity trait), agronomic dates (sowing, anthesis, maturity and harvest dates), and grain yields were recorded for each trial. The detailed information on the locations, field data period, climate, soil properties, and management practices for each experimental station was presented in supplementary data 1. The soil properties including soil texture, bulk density, pH, organic carbon content, and hydraulic properties were obtained from the Global High-Resolution Soil Profile Database (http://dx.doi.org/10.7910/DVN/1PEEY0). The daily weather data on maximum and minimum temperature, precipitation and sunshine hours were extracted from the nearest meteorological station, which were available from the National Meteorological Data Centre (http://data.cma.cn/). The daily solar radiation was calculated by the Angstrom–Prescott equation (Agnström 1924, Prescott 1940).

*2.1.2. Maize phenology and yield observation data*
The observation data on maize phenology and yield were obtained from 121 national agro-meteorological stations, which were maintained by the Chinese Meteorological Administration. The time period for 65 out of 121 stations was from 2005 to 2012, and the rest was consistently from 2010 to 2012. The purpose of these local agro-meteorological stations was to guide local crop production based on meteorological conditions, and provide suggestions for local farmers to manage agricultural meteorological disasters. There were 44 stations in zone I, 49 stations in zone II, and 28 stations in zone III.

*2.1.3. Maize cultivated area*
The dataset of maize cultivation area was obtained from Luo *et al* (2020). They developed a phenology-based classification approach to map three staple crops (maize, wheat, rice) at a resolution of 1 km in China from 2000 to 2015. The accuracy is fairly well with $R^2$ values consistently greater than 0.8 for three crops comparing to the county-level statistical areas. In the current study, maize cultivated area was determined on the basis of the grids with maize cultivation for at least 8 years out of 15 years period.

*2.1.4. Future climate projections data*
Data on future climate projections were dynamically downscaled by the Providing Regional Climates for Impacts Studies based on HadCM3 output. The climate projections over the mainland China were bias-corrected using the observations from 2400 meteorological stations. We separately corrected the four weather variables required by the DSSAT using an additive correction for maximum and minimum

temperature, and a multiplicative adjustment for precipitation and solar radiation (Xie *et al* 2018). The two emission scenarios of RCPs 4.5 and 8.5 were taken into account during two future time periods including the 2030s (2021–2040) and the 2050s (2041–2060), with a baseline of 1986–2005 at a 0.5° × 0.5° resolution.

## 2.2. Methodology

To critically investigate the impacts of projected climate change on maize productivity and identify the time for hybrid adaptation, we conducted the following four procedures (figure 2): (a) calibrating CERES-Maize of DSSAT model based on a large number of field experiments and simulating maize growth and productivity under massive scenarios of realistic ranges of climate, soil, hybrids, and management practices in a given context; (b) developing the RF and XGBoost models based on the output of the above simulations and selecting the hybrid assessment model (i.e. optimal surrogate model) with the highest accuracy for a targeted region; (c) assessing the impacts of climate change on maize productivity using the selected hybrid assessment models combined with the 0.5° × 0.5° grid cells features; and (d) identifying the timescale for hybrid adaptation through threshold-crossing analyses.

### 2.2.1. Calibrating CERES-Maize model and generating pseudo-observations

The DSSAT is one of the most commonly used models that simulate crop growth and productivity considering weather, soil, crop hybrids, and agronomic management practices (Jones *et al* 2003, Hoogenboom *et al* 2019). Maize growth was simulated with the CERES-Maize of DSSAT model using the parameters listed in supplementary table S1. We chose DSSAT from many models mostly because of its advantage in quantifying the roles of agronomic practices in crop growth and productivity, especially hybrids. Moreover, a study on agricultural models intercomparison showed that the outputs from DSSAT were almost near the median of outputs from many models (Rosenzweig *et al* 2014).

In this study, a stepwise calibration was applied by the DSSAT-generalized likelihood uncertainty estimation package—four phenology parameters (P1, P2, P5, PHINT) firstly and then two growth coefficients (G2 and G3). The field data for model calibration and validation were summarized in supplementary table S2. Two metrics, including the root mean square error (RMSE) and relative root mean square error (RRMSE), were used to evaluate model reliability

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

$$RRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i - O_i)^2}}{O_{avg}} \times 100\% \quad (2)$$

where $O_i$ and $S_i$ are the observations and simulations, respectively; $O_{avg}$ is the average values, respectively. $n$ is the number of samples.

To generate a mass of pseudo-observations, we ran CERES-Maize from 1986 to 2005 with six sets of hybrid genetic coefficients which covered the range of currently representative hybrids in terms of climate change adaptation at each experimental station and combined this information with an array of input parameters such as sowing date and planting density that were available from agro-meteorological stations. The sowing dates ranged from the earliest to the latest records with 3 d intervals, and the planting densities were randomly generated within the observed ranges (Zhao and Lobell 2017). In total, at least 30 combinations of input parameters were produced for each run; specifically, 25 200 (30 treatments × 20 years × 7 stations × 6 hybrids) simulations or more were conducted in each region to roughly represent the various management practices and production conditions.

### 2.2.2. Developing hybrid assessment models

To estimate the pixel-based potential impacts of climate change on maize productivity, we coupled CERES-Maize with ML to construct scalable assessment models. Ensemble learning was used to aggregate a collection of algorithms to predict the potential impacts, which represents a better method than that uses any algorithm alone. Here, two ensemble learning algorithms, RF and XGBoost, were trained and validated against the CERES-Maize simulations for the historical (1986–2005) period in each zone. The both algorithms use trees as building blocks to allow for invariance and to scale the inputs and complex interactions of features. The main difference is that the tree in the XGBoost fits on the residual of the former tree to reduce the bias, while RF intends to lower the variance. Their high accuracies and stabilities in agricultural fields have been substantiated in several previous studies, especially for mapping crop planting areas, predicting grain yield, and estimating nitrogen status (Wang *et al* 2019). Details for the two algorithms are given in supplementary text 2.

Many ML methods suffer from accuracy and computational cost of high-dimensional inputs. To reduce features without sacrificing information, we applied Pearson correlation analysis, RF and XGBoost-integrated feature importance to filter out the strongly correlated and non-significant variables. Twenty features were finally used to train the ML models, including five agro-climatic variables, ten topsoil properties, growth duration (DOY), three geographic elements, and carbon dioxide concentration ($CO_2$). The details are presented in table 1 and
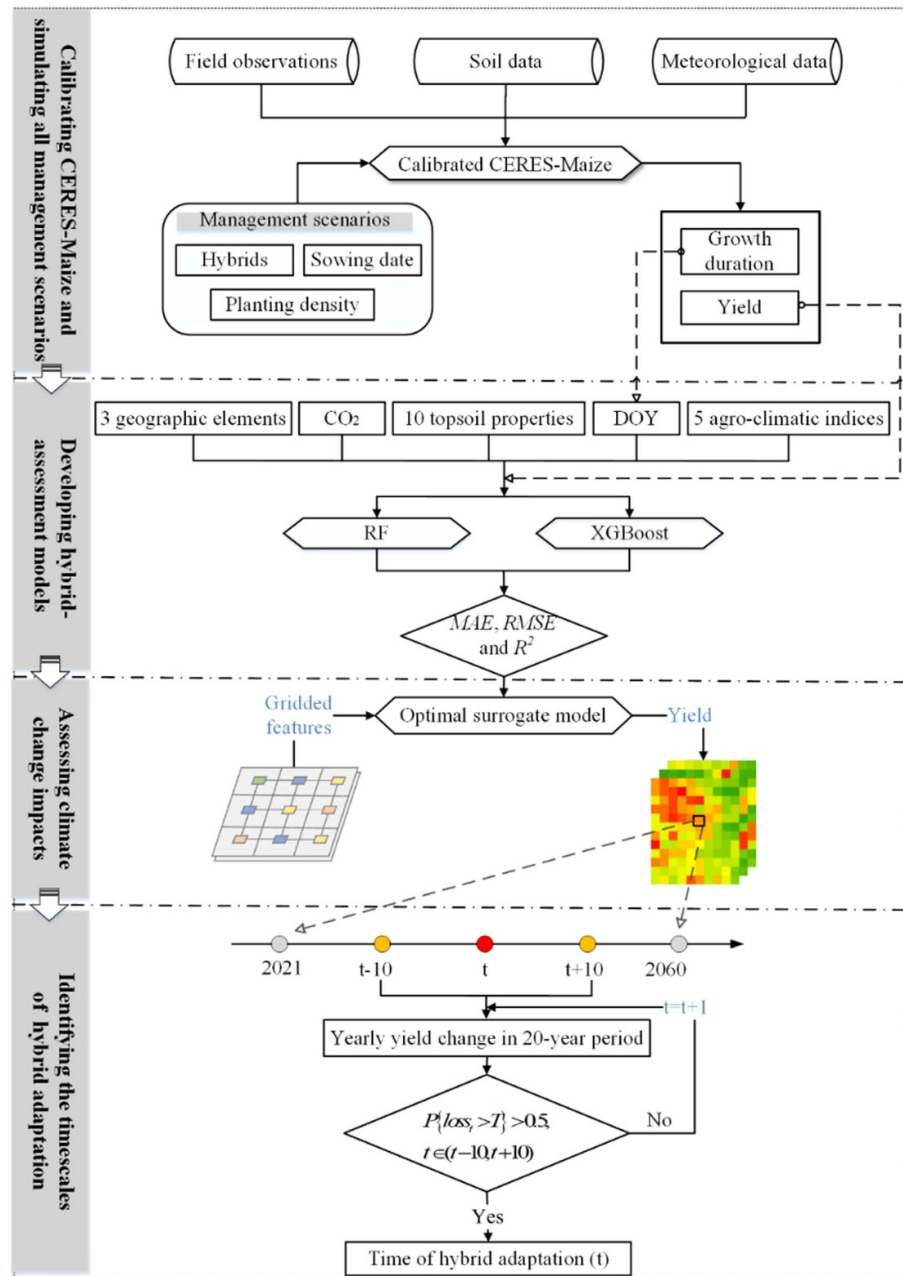
**Figure 2.** Flow chart for assessing the impacts of projected climate change on maize productivity and identifying the timescale of hybrid adaptation.

supplementary table S3. The five agro-climatic variables included growing degree days (GDD), total cold degree days (TCD), frequency of temperatures above 30 °C (OCA), cumulative precipitation (Pgs), and standardized precipitation index (SPI).

The samples were randomly split into 70% for training and 30% for testing, and ten-fold cross-validation and regularization were applied to determine the best hyper-parameters (supplementary table S4). 'Leave-one-out' prediction was conducted to critically evaluate the reliability of the surrogate models. The mean absolute error (MAE), the RMSE, and the $R^2$ were used to evaluate model performance. The optimal surrogate model was the one with the highest $R^2$ and the lowest MAE and RMSE in the test set. We used Python scikit-learn's RF and XGBoost implementation. The $R^2$ and MAE were calculated as follows:

$$R^2 = \frac{\left[\sum_{i=1}^{n}\left(Y_{\text{pred},i} - Y_{\text{avg}}\right)\left(Y_{\text{simu},i} - Y_{\text{avg}}\right)\right]^2}{\sum_{i=1}^{n}\left(Y_{\text{pred},i} - Y_{\text{avg}}\right)^2 \sum_{i=1}^{n}\left(Y_{\text{simu},i} - Y_{\text{avg}}\right)^2} \tag{3}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n}\left|Y_{\text{pred},i} - Y_{\text{simu},i}\right|}{n} \tag{4}$$

where $Y_{\text{pred}}$ is the hybrid model prediction, and $Y_{\text{simu}}$ is the CERES-Maize simulation. $Y_{\text{avg}}$ is the average values. $n$ is the sample size.

**Table 1.** Key features and target variables applied to train ML models.

| Abbreviation | Type | Description | Unit |
|---|---|---|---|
| Agro-climatic indices | Transient[a] | | |
| GDD | | Growing degree days | °Cd |
| TCD | | Total cold degree days | °Cd |
| OCA | | Frequency of temperature above 30 °C | Day |
| Pgs | | Cumulative precipitation | mm |
| SPI | | Standardized precipitation index | — |
| Soil properties | Static | | |
| SLLL | | Lower limit | $cm^3\ cm^{-3}$ |
| SDUL | | Drained upper limit | $cm^3\ cm^{-3}$ |
| SSAT | | Saturated upper limit | $cm^3\ cm^{-3}$ |
| SSKS | | Saturated hydraulic conductivity | $cm\ h^{-1}$ |
| SBDM | | Bulk density | $g\ cm^{-3}$ |
| SLOC | | Organic carbon | % |
| SLCL | | Clay (<0.002 mm) | % |
| SLSI | | Silt (0.05–0.002 mm) | % |
| SLHM | | PH in water | — |
| SCEC | | Cation exchange capacity | $cmol\ kg^{-1}$ |
| Geographic elements | Static | | |
| Latitude | | Grid centroids | Decimal degrees |
| Longitude | | Grid centroids | Decimal degrees |
| Elevation | | Grid averages | m |
| Hybrid variable | Transient | | |
| DOY | | Growth duration | Day |
| Carbon dioxide fertilizer | Transient | | |
| $CO_2$ | | Carbon dioxide concentration | ppm |
| Target variables | Transient | | |
| Yield | | Maize yield | $kg\ ha^{-1}$ |

[a] Transient type indicates a variable that changes with time and static one is the opposite.

A critical issue for evaluating the impacts of climate change was the reliability of the surrogate models. To address this concern, we further evaluated the surrogate models against the actual measurements of maize phenology and yield obtained from 121 agrometeorological stations. In total, the surrogate models' estimates were compared to the observed yields of 688 unique observations from various years and locations across the Chinese maize belt.

*2.2.3. Assessing climate change impacts on maize development and productivity*
To quantify the impacts of climate change on maize productivity, we first simulated annual phenology of each hybrid at each experimental station in the baseline period (1986–2005), 2030s (2021–2040), and 2050s (2041–2060) under RCP 4.5 and RCP 8.5, respectively. Agronomic management practices (i.e. sowing date, planting density, row spacing, irrigation and fertilization applications) were fixed at the baseline levels, representing no-adaptation scenario. $CO_2$ concentration was set at 380 ppm in the baseline period (Deryng *et al* 2016), 447 (499) ppm and 470 (571) ppm in the 2030s (2050s) under RCP 4.5 and RCP 8.5, respectively (IPCC 2014). And then, we upscaled the simulated phenology to grid cells

with a 0.5° × 0.5° resolution using kriging method in ArcGIS 10.6 software and calculated gridded agroclimatic variables. Finally, we applied the surrogate models in three periods: baseline, 2030s, and 2050s under the two RCPs (4.5 and 8.5) scenarios. We compared the annual mean yields of a specific hybrid for each scenario at the pixel scale with their levels during baseline period. The changes in yield were quantified using equation (5):

$$Y_c = \frac{Y_f - Y_b}{Y_b} \times 100\% \qquad (5)$$

where $Y_c$ is the projected yield change, and $Y_f$ and $Y_b$ are the annual mean yield in the future and baseline periods, respectively.

*2.2.4. Identifying the timescales of hybrid adaptation*
Threshold-crossing analyses have been widely used in climate-related studies, especially for evidence-based adaptation (Rippke *et al* 2016). To critically investigate when and where hybrid adaptation is required, a 5%, 10% and 15% yield loss value are set as the suitability threshold for hybrid renewal, respectively (Lobell *et al* 2008). A potential adaptation will be required for the location when yield loss is greater than the suitability threshold in more

than 10 years during the 20 years running period. Such a period is generally accepted as a suitable time window because 2 decades adequately reflect the progressive changes in climate and the adaptability state (Rippke *et al* 2016). Moreover, a 50% risk is a trade-off among diverse losses in broad cultivation areas and complex environments. Here, we characterized the mean, earliest and latest times for hybrid adaptation, which corresponded to the median, maximum and minimum losses of the six hybrids, respectively.

## 3. Results

### 3.1. Reliability of the hybrid assessment models

The statistical metrics between the simulated yields by CERES-Maize (see supplementary text 3 and figure S1 for the calibration and validation of CERES-Maize model) and the predicted yields by RF and XGBoost in each agro-ecological zone are summarized in table 2. Both models successfully explain more than 85% of the simulated maize yield variance, with the bias RRMSE less than 7% and RMSE < 940 kg ha$^{-1}$. The accuracies of the hybrid models are higher than those reported in previous studies that coupled crop model and linear regression to predict crop yield (Lobell *et al* 2015, Burke and Lobell 2017, Feng *et al* 2019, 2020). The MAEs range from 516 to 806 kg ha$^{-1}$ or 3.8% to 6.0% of the mean simulated yields (13 499 kg ha$^{-1}$) across the agro-ecological zones. The probability distributions of yields derived by the two ML models are generally in agreement with those simulated by CERES-Maize, except slight underestimation of nadirs and overestimation of apexes supplementary figure S2.

The surrogate models could predict fairly well the observed yields in the 688 observations with all combinations of years and locations obtained from 121 agro-meteorological stations across the Chinese maize belt (figure 3), with $R^2$ ranging from 0.45 to 0.50 and 0.49 to 0.54 for RF and XGBoost, respectively. The yields derived by the both models were gathered around the 1:1 line (figure 3) and the forecast errors showed normal distribution, indicating non-significant bias estimations on the whole (supplementary figure S3). Yet we noticed that the two models underestimated maize yield in zone I when the observations were above 13 000 kg ha$^{-1}$. Such bias may be that the ML models were trained by six common hybrids with multi-scenario simulations, but the agro-meteorological observations included extremely high yields from superior hybrids adopted by the farmers in a specific year. When combined into a single comparison of all the 688 observations, the $R^2$ reached up to 0.48 for RF model and 0.56 for XGBoost
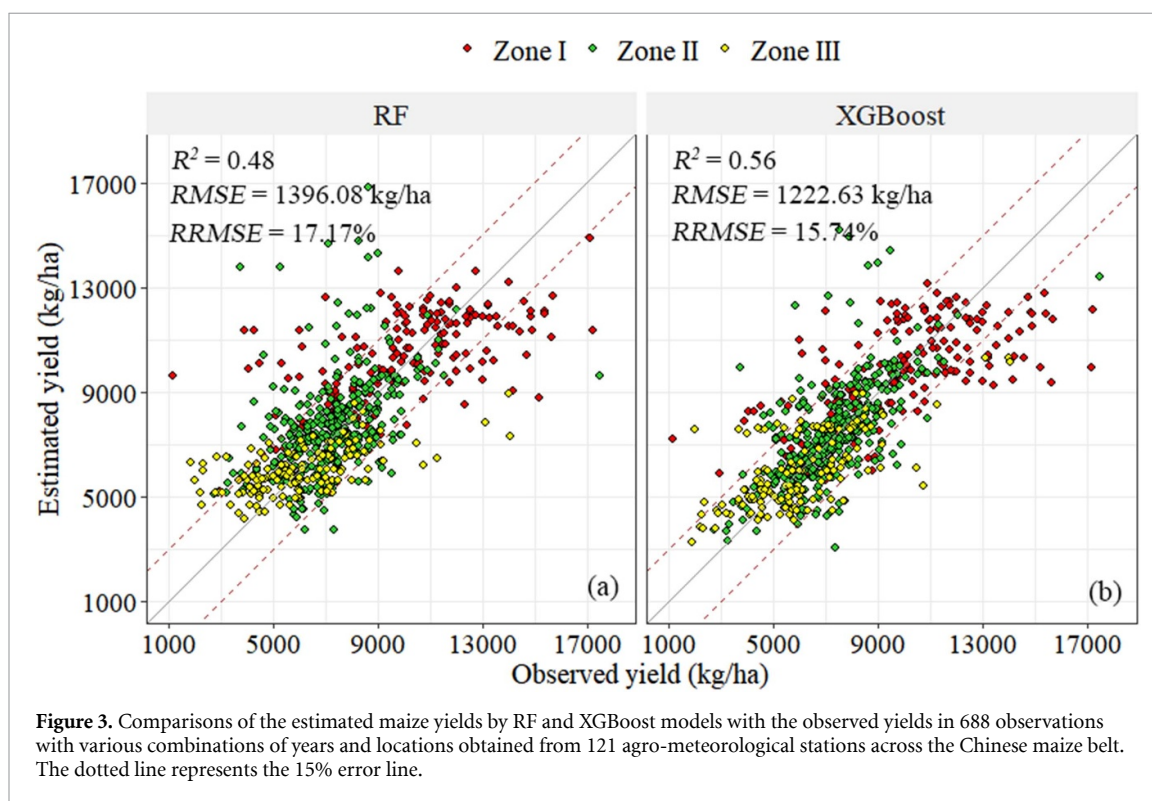
model. The $R^2$ was slightly higher than those previously obtained with a similar hybrid scheme or crop model assimilation method (Huang *et al* 2015, Lobell *et al* 2015, Burke and Lobell 2017, Chen *et al* 2018a). Namely, the surrogate models capture significantly about half of the actual yield variations across the Chinese maize belt and can be acceptable to assess the impacts of climate change at a regional scale. As a whole, XGBoost was more robust than RF as XGBoost had a higher accuracy and more identical fitting curves (figure 3 and supplementary figure S2). Therefore, the XGBoost models were applied to assess the climate impacts at the scale of agro-ecological zone. The relative importance of the 20 features from the XGBoost surrogate models was given in supplementary text 4 and figure S4.

### 3.2. Impacts of projected climate change on maize productivity for different hybrids

The results showed that the projected climate change (see supplementary text 5 and supplementary figure S5 for the projected changes in agro-meteorological indices) would mostly have negative impacts on maize yields throughout the cultivation areas under the both scenarios. The impact magnitudes would differ remarkably among the cultivated hybrids. The largest yield loss was expected in zone I, which was projected to be nearly 6.8% and 6.4% in the 2030s under RCP 4.5 and RCP 8.5, respectively (figures 4(a1) and (a3)). It would reach up to approximately 10% in the 2050s under the both scenarios (figures 4(a2) and (a4)). In addition, the most distinct divergences in the responses of different hybrids to climate change would be in this zone. For example, yield loss would be 2.3% in the 2030s under RCP 4.5 for late-maturing hybrids (L2), but would likely be tripled (7.1%) for early-maturing hybrids (E1) (figure 4(a1)). In comparison with zone I, yield loss in zone II would be relatively smaller but could have a larger spatial variability as illustrated by the longer whiskers of each box plot. In addition, it would be highly dependent on the time periods. Yield loss would be nearly twice as high in the 2050s (7.9% and 6.0% under RCP 4.5 and RCP 8.5, respectively) (figures 4(b2) and (b4)) as in the 2030s (3.4% and 2.9%) (figures 4(b1) and (b3)). Climate change would have the smallest impacts on maize yields in zone III, where yield loss would be consistently lower than 2% except for a late-maturing hybrid (L2) (figures 4(c1)–(c4)). For the whole country, without adaptation, maize yield would decrease by 5.8% (4.7%) and 8.5% (7.8%) in the 2030s (2050s) under RCP 4.5 and RCP 8.5, respectively (figures 4(d1)–(d4)). More details on the spatial distributions and interannual variability of the projected yield changes are presented in the supplementary text 6 and text 7.

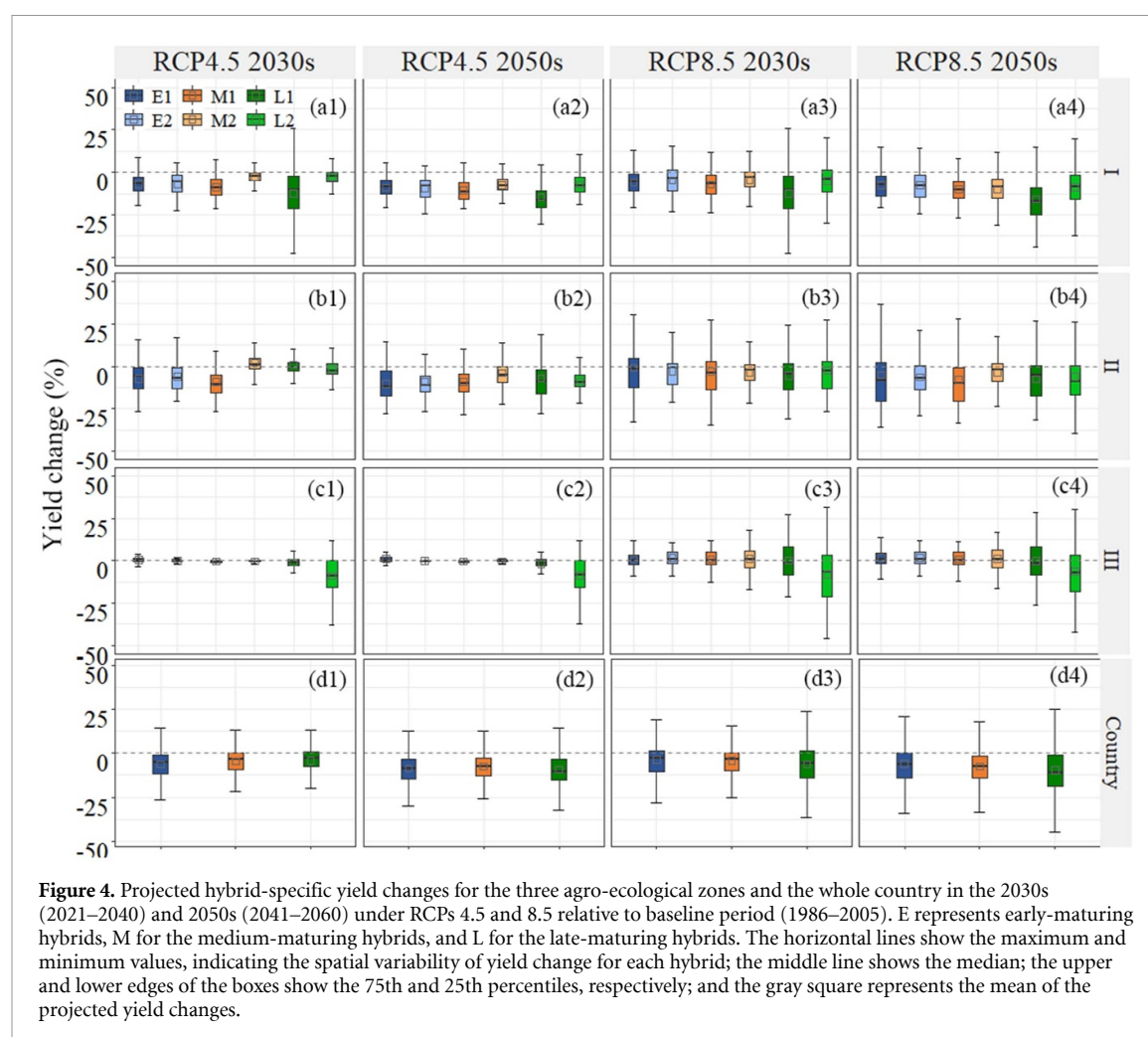**Table 2.** Performance of RF and XGBoost models in predicting maize yields in the three agro-ecological zones.

| Zone | RF | | | XGBoost | | |
|------|------|------|-------|------|------|-------|
| | MAE | RMSE | $R^2$ | MAE | RMSE | $R^2$ |
| I | 683.18 | 961.02 | 0.86 | 634.76 | 853.45 | 0.87 |
| II | 806.12 | 1065.26 | 0.90 | 654.25 | 910.90 | 0.92 |
| III | 586.94 | 793.49 | 0.85 | 516.01 | 644.17 | 0.88 |
| Average | 692.08 | 939.92 | 0.87 | 601.67 | 802.84 | 0.89 |



**Figure 3.** Comparisons of the estimated maize yields by RF and XGBoost models with the observed yields in 688 observations with various combinations of years and locations obtained from 121 agro-meteorological stations across the Chinese maize belt. The dotted line represents the 15% error line.

### 3.3. Timescale for hybrid adaptation

Hybrid adaptation would likely be required in most of the cultivation areas under RCP 4.5 and RCP 8.5 with the suitability threshold of 10%, although there would be remarkable differences among the locations and times (figures 5(c) and (d)). More than half of the cultivated grids (54%) would require hybrid adaptation by 2050 under RCP 4.5 (figure 5(c)). This value would maintain at 50%, but the time would become earlier—by 2040—under RCP 8.5 (figure 5(d)). Interestingly, the areas that would require the most urgent hybrid adaptation were projected to be in Northeast China under RCP 4.5, where improved hybrids should be in place before 2040 for more than 73% of the grids (figure 5(c)). Similar results were found in zone II under RCP 8.5 for approximately 58% of the grids (figure 5(d)). By contrast, less than 10% of the grids in zone III would exceed the suitability threshold by 2050 under the both scenarios (figures 5(c) and (d)). In most areas, the time of hybrid adaptation would differ remarkably among the two RCPs. For example, genetic adaptation would be needed a decade later under RCP 4.5 than under RCP 8.5 in zone II, suggesting that mitigating greenhouse gas emissions could allow more time for preparing adaptations (figures 5(c) and (d)). In addition, the times requiring hybrid adaptation would vary among agro-ecological zones under the two scenarios (figures 5(c) and (d)). To critically investigate the timing of hybrid adaptation, the threshold of 5% (figures 5(a) and (b)) and 15% (figures 5(e) and (f)) were also examined. We found that approximately 21% of cultivated areas would require hybrid renewal before 2050 under the both scenarios with the threshold of 15% (figures 5(e) and (f)). The situation would be more severe for risk-averse farmers (the threshold of 5%) since the existing hybrids will no longer be suitable for 69% of the cultivated grids (figures 5(a) and (b)). Breeding a new crop hybrid may generally take about 20 years or more (Beyene and Kassie 2015, Challinor *et al* 2016, Cowling *et al* 2018). Therefore, breeding investment should be started at once because there is only a 20 years time window left from now to the time when most areas require hybrid adaptation. The earliest and latest times of hybrid adaptation for each suitability
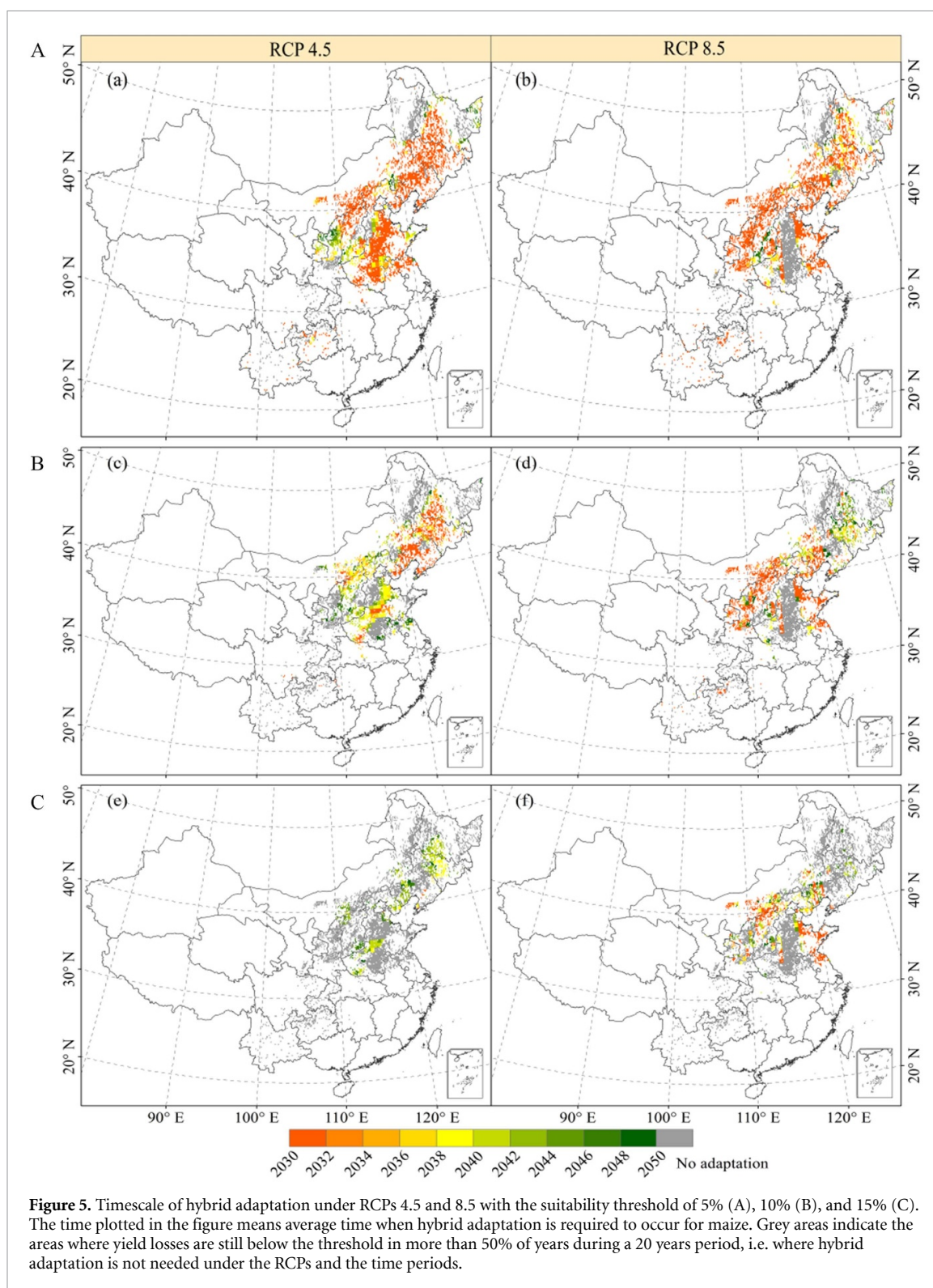
**Figure 4.** Projected hybrid-specific yield changes for the three agro-ecological zones and the whole country in the 2030s (2021–2040) and 2050s (2041–2060) under RCPs 4.5 and 8.5 relative to baseline period (1986–2005). E represents early-maturing hybrids, M for the medium-maturing hybrids, and L for the late-maturing hybrids. The horizontal lines show the maximum and minimum values, indicating the spatial variability of yield change for each hybrid; the middle line shows the median; the upper and lower edges of the boxes show the 75th and 25th percentiles, respectively; and the gray square represents the mean of the projected yield changes.

threshold are summarized in the supplementary text 8 and figure S8.

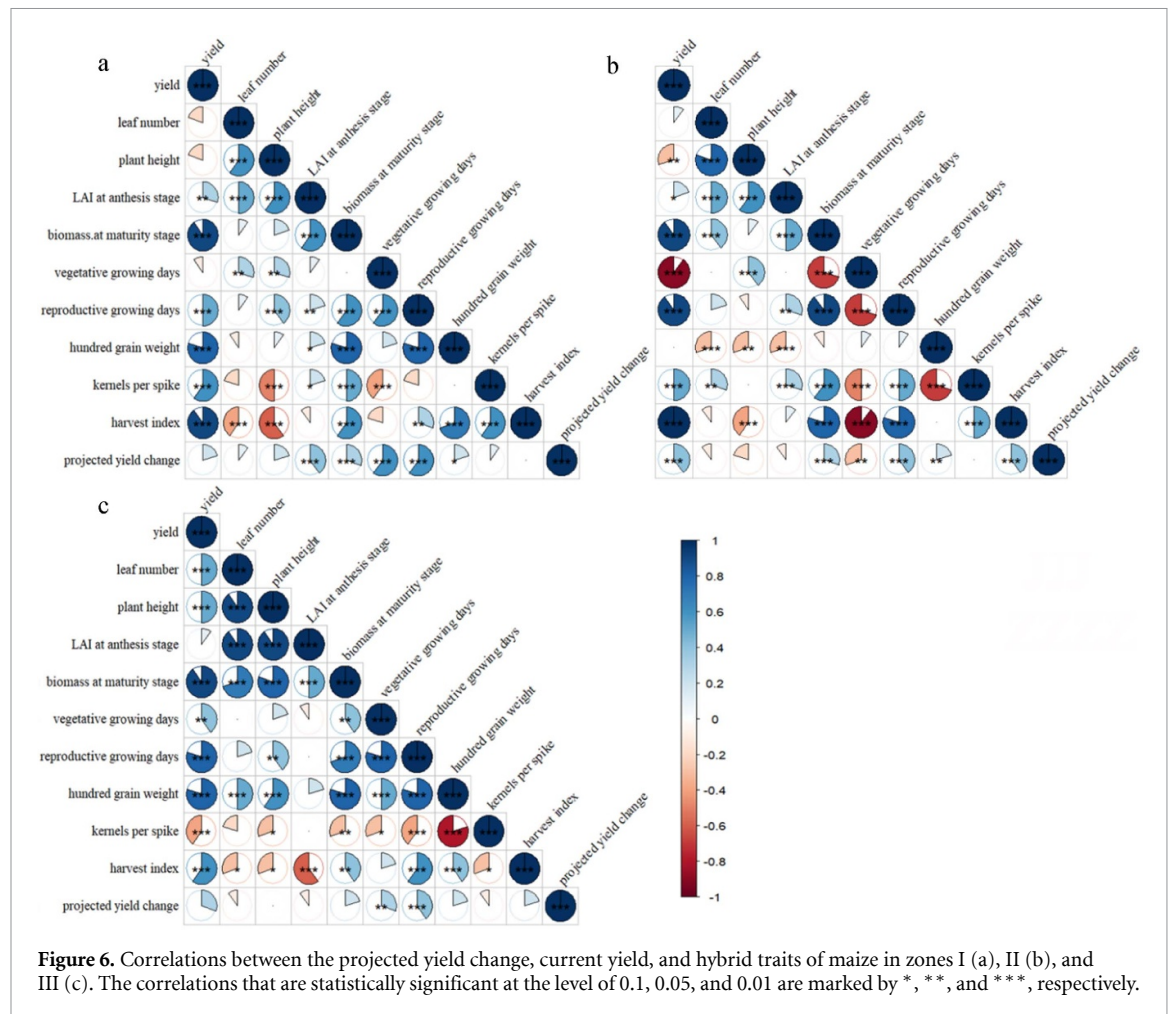### 3.4. The hybrid ideotypic traits for adapting to climate change

To further identify the hybrid ideotypic traits for adapting to climate change, we calculated the correlations between the projected yield changes, current yield, and hybrid traits based on the multi-scenario simulations of CERES-Maize model for the six typical hybrids in each of the three agro-ecological zones (figure 6). Various plant traits of the 18 hybrids were given in supplementary data 2. The reliability of the correlations was validated by conducting the correlation analyses on the available eight-year field experiments data for the hybrids XY335 (M2) and ZD958 (L2) in zone I (supplementary text 9 and figure S9). In zone I, the projected yield loss was significantly and positively correlated with leaf area index (LAI) at anthesis, biomass at maturity, vegetative and reproductive growth durations ($p < 0.01$), and hundred-grain weight ($p < 0.1$) (figure 6(a) and supplementary figure S9). In other words, the hybrids with high values for these traits would have less yield loss under

future climate change. Combining the changes in projected yield (figure 4) and growth duration (supplementary figure S10), the breeding efforts should aim to breed the medium-maturing hybrids with a large LAI (high canopy light interception), biomass (high light use efficiency) and grain weight (high grain-filling rate) for zone I. In comparison with zone I, the projected yield loss in zone II had positive correlations with current yield, biomass, reproductive growth duration, hundred-grain weight, and harvest index ($p < 0.05$), but negatively correlated with LAI and vegetative growth duration (figure 6(b)). Therefore, the breeding efforts should aim to reduce the ratio of vegetative to reproductive growth duration (increasing grain-filling duration), reduce LAI (reducing water consumption), and increase biomass (increasing light use efficiency) for zone II. In zone III, growth duration was the most important trait that affected the projected yield changes (figure 6(c)). The medium- or late-maturing hybrids could maintain or even slightly increase yield under RCP 4.5 and 8.5, taking advantage of the ameliorating photothermal and water resources with climate change in the zone. The correlations are statistically significant ($p < 0.1$),

**Figure 5.** Timescale of hybrid adaptation under RCPs 4.5 and 8.5 with the suitability threshold of 5% (A), 10% (B), and 15% (C). The time plotted in the figure means average time when hybrid adaptation is required to occur for maize. Grey areas indicate the areas where yield losses are still below the threshold in more than 50% of years during a 20 years period, i.e. where hybrid adaptation is not needed under the RCPs and the time periods.

the identified ideotypic traits have sound biophysical and physiological basis and are insightful for potential long-term breeding, although the genetic correlations might change over the course of recurrent selection. As a whole, the medium-maturing maize hybrids with a long grain-filling duration and a high light use efficiency could be climate resilient hybrids and

breeding efforts to adapt to climate change, although the ideotypic traits could be different for a specific environment. The results are supported by previous studies that design ideotypes for different crops and cultivation environments (Harrison *et al* 2014, Tao *et al* 2017, Cowling *et al* 2018, Asseng *et al* 2019, Xiao *et al* 2020).

**Figure 6.** Correlations between the projected yield change, current yield, and hybrid traits of maize in zones I (a), II (b), and III (c). The correlations that are statistically significant at the level of 0.1, 0.05, and 0.01 are marked by *, **, and ***, respectively.

## 4. Discussion

We showed that the impacts of climate change were mostly negative throughout the maize cultivation areas due to decrease in crop growth duration (supplementary text 10 and figure S10). The results were supported by some previous studies based on crop models driven by multiple future climate scenarios (Chen *et al* 2018b) and empirical analyses in China (Zhang *et al* 2016, 2018). Furthermore, we found that the spatial patterns and magnitudes of projected yield changes would depend on maize hybrids. For example, in Northeast China, maize yield would decrease by more than 10% in approximately 84% of the grids with early-maturing hybrids, but for the late hybrids-maturing it would decrease by less than 5% in 21% of the grids, in the 2030s under RCP 4.5 (supplementary figures S6(a1) and (a3)). That was, the late-maturing hybrids did not consistently suffer from greater yield losses under the climate change, suggesting that other traits might be at work such as heat- and drought-resistance, grain-filling rate and light use efficiency. Our findings indicated that linking yield loss to temperature increases without considering hybrids adaptability may cause biased estimations

on crop response to climate change (Ray *et al* 2015, Shew *et al* 2020).

In absence of effective adaptations, more than half of the cultivated areas would require hybrid adaptation by 2050 under the both scenarios with the suitability threshold of 10% (medium risk). The sensitive areas cover $2.16 \times 10^3$ Mha, mostly located in Northeast China and the North China Plain, which account for 56% of national total maize production. Given that most commonly used hybrids across the main cultivation areas of China have been investigated in this study, we argued that breeding actions should be triggered immediately in China, as several studies have suggested for other areas (Challinor *et al* 2016, Cowling *et al* 2018).

The ideotypic traits identified for each of the three agro-ecological zones have sound biophysical and physiological basis. In zone I, the photothermal and water resources will be ameliorated much by climate change (figure 6(a) and supplementary figure S5), which will provide a great potential for the hybrids with a relatively long growth duration and large LAI (Liu *et al* 2013). In zone II, the photothermal resources will be ameliorated moderately by climate change but water stress may be aggravated

(figure 6(b) and supplementary figure S5), so the hybrids with short vegetative growth duration and less LAI will be promising (Tao *et al* 2015, Kromdijk *et al* 2016, de Souza and Long 2018). In zone III, the photothermal resources will be ameliorated by climate change (figure 6(c) and supplementary figure S5), the medium- or late-maturing hybrids will take advantage of the ameliorated photothermal resources in the zone (Chen *et al* 2018b). Considering the time lag from the development of improved hybrids to the seed ready for farmers use, we suggest that breeding action should be triggered immediately to maintain a stable maize production by 2050 in China. In addition to the hybrid adaptation, two alternative adaptations are applicable: transformational adaptation and optimizing G × E × M interactions. For example, shifting to more heat- and drought-tolerant cereals, such as millet and sorghum, would offset some of the yield losses (Park *et al* 2012). Adopting optimal G × E × M interaction strategy for a target environment could narrow yield gap and buy extra time for breeding (Zimmermann *et al* 2017).

Extensive genetic and physiological studies on crop hybrids have been conducted which are essential to produce high and stable grain yield. Nevertheless, interdisciplinary studies are necessary to develop climate resilient crop hybrids for the future. Our study demonstrates a novel and sound interdisciplinary approach to plan hybrid adaptation for a changing climate based on crop modelling and big data, which can provide useful information for hybrid breeding and climate change adaptation. This study, like many others, also has some uncertainties. First, the climate change scenarios and subsequently the estimated impacts could have some uncertainties since the climate models suffer from inherent systematic bias (Glotter *et al* 2014) and the assessment is based on single climate model (HadCM3) output. To acquire more reliable results, an ensemble of fine-spatial-scale climate projections should be applied. Second, several studies have reported that the uncertainty from crop model structure was larger than that from downscaled GCM projections (Asseng *et al* 2013, Maraun *et al* 2017, Zhang *et al* 2017). Selection of crop model and ML model could bring uncertainties. Because the existing process-based models apply different functions to represent the interactions between crop and weather factors (Zhang *et al* 2017, Tao *et al* 2018), multi-model simulations could obtain more robust estimations. Likewise, selecting the most appropriate method is vital for a specific study due to various structures of ML algorithms. Here, we carefully calibrated the CERES-Maize model with more than 70 field experiments and assessed the impacts of projected climate change using an ensemble of hybrid-specific estimations, which might address these uncertainties to some extent. Third, the time for hybrid adaptation

might have some uncertainties because we remained other agronomic managements such as sowing date and planting density constant in the future and did not explicitly take into account the influences of other extreme climate events, pests and diseases, as well as socio-economic constraints. In addition, the adaptation time might be conservative in zone I because the surrogate models had downward bias for the observed yield and consequently underestimated the climate change impacts. Finally, we separately tested several key phenotypic traits related to canopy structure and photosynthetic efficiency, the genetic and physiological basis of the ideotypic traits and the performance of combined traits should be further investigated through detailed biological researches. Despite of these limitations, we critically investigated the impacts of climate change on maize productivity using a novel approach based on a large number of field hybrids trials, systematically analysed the adaptability of existing hybrids, and identified the locations, timescale, and ideotypic traits for hybrid adaptation. Our findings highlight the necessity and urgency of breeding climate resilient hybrids, providing policy-makers with the early signals of when and where the adaptability of current hybrids would be broken, and stimulating proactive investment to facilitate breeding.

## 5. Conclusions

Using a novel hybrid assessment model that coupled CERES-Maize with ML, we critically investigated the impacts of climate change on maize productivity with an ensemble of hybrid-specific estimations in three agro-ecological zones across China, and critically identified the timescale of hybrid adaptation with a series of suitability thresholds. The results showed that climate change would have mostly negative impacts on maize yields in the cultivation areas under the both scenarios, and the magnitudes of yield reductions would highly depend on the growth cycle of the cultivated hybrids. Current hybrids replacement could only alleviate but not offset the yield loss due to climate change. Without adaptations, national production would suffer from a 7.6% decrease under the both scenarios. Threshold-crossing analyses suggested that approximately 53% of the cultivation areas would require breeding intervention before 2050 under both the emission pathways. The hybrids with a medium growth cycle, long grain-filling period, and high photosynthetic capacity should be breeding targets to adapt to climate change, although the ideotypic traits could be different for a specific environment. Only 2 decades of leading time remain in Northeast China and the Huang-Huai-Hai Plain. Our findings provide the time and location, as well as the hybrid ideotypic traits, for maize production to precisely adapt to climate change.

## Data availability statement

## Code availability

## Acknowledgments

## Conflict of interest

The authors declare no competing interests.

## References

Agnström A 1924 Solar and terrestrial radiation.19 *Mon. Weather Rev.* **52** 83

Asseng S *et al* 2013 Uncertainty in simulating wheat yields under climate change *Nat. Clim. Change* **3** 827–32

Asseng S *et al* 2019 Climate change impact and adaptation for wheat protein *Glob. Change Biol.* **25** 155–73

Azzari G, Jain M and Lobell D B 2017 Towards fine resolution global maps of crop yields: testing multiple methods and satellites in three countries *Remote Sens. Environ.* **202** 129–41

Bailey-Serres J, Parker J E, Ainsworth E A, Oldroyd G E D and Schroeder J I 2019 Genetic strategies for improving crop yields *Nature* **575** 109–18

Beyene A D and Kassie M 2015 Speed of adoption of improved maize varieties in Tanzania: an application of duration analysis *Technol. Forecast. Soc. Change* **96** 298–307

Burke M and Lobell D B 2017 Satellite-based assessment of yield variation and its determinants in smallholder African systems *Proc. Natl Acad. Sci. USA* **114** 2189–94

Challinor A J, Koehler A K, Ramirez-Villegas J, Whitfield S and Das B 2016 Current warming will reduce yields unless maize breeding and seed systems adapt immediately *Nat. Clim. Change* **6** 954–8

Chen Y, Zhang Z and Tao F 2018a Improving regional winter wheat yield estimation through assimilation of phenology and leaf area index from remote sensing data *Eur. J. Agron.* **101** 163–73

Chen Y, Zhang Z and Tao F 2018b Impacts of climate change and climate extremes on major crops productivity in China at a global warming of 1.5 °C and 2.0 °C *Earth Syst. Dyn.* **9** 543–62

Cowling W A, Li L, Siddique K H M, Banks R G and Kinghorn B P 2018 Modeling crop breeding for global food security during climate change *Food Energy Secur.* **8** e00157

Crane-Droesch A 2018 Machine learning methods for crop yield prediction and climate change impact assessment in agriculture *Environ. Res. Lett.* **13** 114003

de Souza A P and Long S P 2018 Toward improving photosynthesis in cassava: characterizing photosynthetic limitations in four current African cultivars *Food Energy Secur.* **7** e00130

Deryng D *et al* 2016 Regional disparities in the beneficial effects of rising $CO_2$ concentrations on crop water productivity *Nat. Clim. Change* **6** 786–90

Faostat F 2016 *FAOSTAT Statistical Database* (Rome: FAO (Food and Agriculture Organization of the United Nations))

Feng P *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** 107922

Feng P, Wang B, Liu D L, Waters C and 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** 100–13

Folberth C *et al* 2019 Spatio-temporal downscaling of gridded crop model yield estimates based on machine learning *Agric. For. Meteorol.* **264** 1–15

Glotter M *et al* 2014 Evaluating the utility of dynamical downscaling in agricultural impacts projections *Proc. Natl Acad. Sci. USA* **111** 8776–81

Grimm R, Behrens T, Märker M and Elsenbeer H 2008 Soil organic carbon concentrations and stocks on Barro Colorado Island—digital soil mapping using random forests analysis *Geoderma* **146** 102–13

Harrison M T, Tardieu F, Dong Z, Messina C D and Hammer G L 2014 Characterizing drought stress and trait influence on maize yield under current and future conditions *Glob. Change Biol.* **20** 867–78

Hoogenboom G *et al* 2019 Decision support system for agrotechnology transfer (DSSAT) version 4.7.5 (Gainesville, FL: DSSAT Foundation) (available at: http://dssat.net)

Huang J *et al* 2015 Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model *Agric. For. Meteorol.* **204** 106–21

IPCC 2014 Climate change 2014: synthesis report *Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* ed Core Writing Team, R K Pachauri and L A Meyer (Geneva: IPCC) p 151

Jones J W *et al* 2003 The DSSAT cropping system model *Eur. J. Agron.* **18** 235–65

Kromdijk J *et al* 2016 Improving photosynthesis and crop productivity by accelerating recovery from photoprotection *Science* **354** 857–61

Lesk C, Rowhani P and Ramankutty N 2016 Influence of extreme weather disasters on global crop production *Nature* **529** 84–87

Li S, Tompkins A M, Lin E and Ju H 2016 Simulating the impact of flooding on wheat yield—case study in East China *Agric. For. Meteorol.* **216** 221–31

Liu Z, Hubbard K G, Lin X and Yang X 2013 Negative effects of climate warming on maize yield are reversed by the changing of sowing date and cultivar selection in Northeast China *Glob. Change Biol.* **19** 3481–92

Lobell D B *et al* 2008 Prioritizing climate change adaptation needs for food security in 2030 *Science* **319** 607–10

Lobell D B, Thau D, Seifert C, Engle E and Little B 2015 A scalable satellite-based crop yield mapper *Remote Sens. Environ.* **164** 324–33

Luo Y, Zhang Z, Chen Y, Li Z and Tao F 2020 ChinaCropPhen1km: a high-resolution crop phenological dataset for three staple crops in China during 2000–2015 based on leaf area index (LAI) products *Earth Syst. Sci. Data* **12** 197–214

Maraun D *et al* 2017 Towards process-informed bias correction of climate change simulations *Nat. Clim. Change* **7** 764–73

Martre P *et al* 2015 Multimodel ensembles of wheat growth: many models are better than one *Glob. Change Biol.* **21** 911–25

Morris M, Dreher K, Ribaut J-M and Khairallah M 2003 Money matters (II): costs of maize inbred line conversion schemes at CIMMYT using conventional and marker-assisted selection *Mol. Breed.* **11** 235–47

Muller C *et al* 2019 The global gridded crop model intercomparison phase 1 simulation dataset *Sci. Data* **6** 50

Parent B *et al* 2018 Maize yields over Europe may increase in spite of climate change, with an appropriate use of the genetic variability of flowering time *Proc. Natl Acad. Sci. USA* **115** 10642–7

Park S E *et al* 2012 Informing adaptation responses to climate change through theories of transformation *Glob. Environ. Change* **22** 115–26

Prescott J 1940 Evaporation from a water surface in relation to solar radiation *Trans. R. Soc. South Aust.* **64** 114–8

Ramirez-Villegas J *et al* 2018 Breeding implications of drought stress under future climate for upland rice in Brazil *Glob. Change Biol.* **24** 2035–50

Ray D K, Gerber J S, MacDonald G K and West P C 2015 Climate variation explains a third of global crop yield variability *Nat. Commun.* **6** 5989

Ray D K, Ramankutty N, Mueller N D, West P C and Foley J A 2012 Recent patterns of crop yield growth and stagnation *Nat. Commun.* **3** 1293

Reichstein M *et al* 2019 Deep learning and process understanding for data-driven Earth system science *Nature* **566** 195–204

Rippke U *et al* 2016 Timescales of transformational climate change adaptation in sub-Saharan African agriculture *Nat. Clim. Change* **6** 605–9

Rosenzweig 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** 3268–73

Shew A M, Tack J B, Nalley L L and Chaminuka P 2020 Yield reduction under climate warming varies among wheat cultivars in South Africa *Nat. Commun.* **11** 4408

Stevens T and Madani K 2016 Future climate impacts on maize farming and food security in Malawi *Sci. Rep.* **6** 36241

Tao F *et al* 2017 Designing future barley ideotypes using a crop model ensemble *Eur. J. Agron.* **82** 144–62

Tao F *et al* 2018 Contribution of crop model structure, parameters and climate projections to uncertainty in climate change impact assessments *Glob. Change Biol.* **24** 1291–307

Tao F, Zhang Z, Liu J and Yokozawa M 2009 Modelling the impacts of weather and climate variability on crop productivity over a large area: a new super-ensemble-based probabilistic projection *Agric. For. Meteorol.* **149** 1266–78

Tao F, Zhang Z, Zhang S and Rötter R P 2015 Heat stress impacts on wheat growth and yield were reduced in the Huang-Huai-Hai plain of China in the past three decades *Eur. J. Agron.* **71** 44–52

Wang S, Azzari G and Lobell D B 2019 Crop type mapping without field-level labels: random forest transfer and unsupervised clustering techniques *Remote Sens. Environ.* **222** 303–17

Xiao D, Liu D L, Wang B, Feng P and Waters C 2020 Designing high-yielding maize ideotypes to adapt changing climate in the North China Plain *Agric. Syst.* **181** 102805

Xie W *et al* 2018 Decreases in global beer supply due to extreme drought and heat *Nat. Plants* **4** 964–73

Yang Y *et al* 2019 Improving maize grain yield by matching maize growth and solar radiation *Sci. Rep.* **9** 3635

Zhang L, Zhang Z, Chen Y, Wei X and Song X 2018 Exposure, vulnerability, and adaptation of major maize-growing areas to extreme temperature *Nat. Hazards* **91** 1257–72

Zhang S, Tao F and Zhang Z 2017 Uncertainty from model structure is larger than that from model parameters in simulating rice phenology in China *Eur. J. Agron.* **87** 30–39

Zhang Y, Zhao Y, Wang C and Chen S 2016 Using statistical model to simulate the impact of climate change on maize yield with climate and crop uncertainties *Theor. Appl. Climatol.* **130** 1065–71

Zhao C *et al* 2017 Temperature increase reduces global yields of major crops in four independent estimates *Proc. Natl Acad. Sci. USA* **114** 9326–31

Zhao Y and Lobell D B 2017 Assessing the heterogeneity and persistence of farmers' maize yield performance across the North China Plain *Field Crops Res.* **205** 55–66

Zheng D *et al* 2018 Effect of freezing temperature and duration on winter survival and grain yield of winter wheat *Agric. For. Meteorol.* **260–261** 1–8

Zimmermann A *et al* 2017 Climate change impacts on crop yields, land use and environment in response to crop sowing dates and thermal time requirements *Agric. Syst.* **157** 81–92