





Article

The Prediction of Wheat Yield in the North China Plain by Coupling Crop Model with Machine Learning Algorithms

Yanxi Zhao ^{1,2,3} , Dengpan Xiao ^{1,2,3,*} , Huizi Bai ^{1,*}, Jianzhao Tang ¹, De Li Liu ^{4,5} , Yongqing Qi ⁶ 
and Yanjun Shen ^{6,7}

- ¹ Hebei Technology Innovation Center for Geographic Information Application, Institute of Geographical Sciences, Hebei Academy of Sciences, Shijiazhuang 050011, China
 - ² College of Geography Science, Hebei Normal University, Shijiazhuang 050024, China
 - ³ Hebei Laboratory of Environmental Evolution and Ecological Construction, Shijiazhuang 050024, China
 - ⁴ NSW Department of Primary Industries, Wagga Wagga Agricultural Institute, Wagga, NSW 2650, Australia
 - ⁵ Climate Change Research Centre, University of New South Wales, Sydney, NSW 2052, Australia
 - ⁶ Key Laboratory for Agricultural Water Resources, Hebei Key Laboratory for Agricultural Water Saving, Center for Agricultural Resources Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Shijiazhuang 050021, China
 - ⁷ School of Advanced Agricultural Sciences, University of the Chinese Academy of Sciences, Beijing 100049, China
- * Correspondence: xiaodp@sjziam.ac.cn (D.X.); baihz272@126.com (H.B.)

Abstract: The accuracy prediction for the crop yield is conducive to the food security in regions and/or nations. To some extent, the prediction model for crop yields combining the crop mechanism model with statistical regression model (SRM) can improve the timeliness and robustness of the final yield prediction. In this study, the accumulated biomass (AB) simulated by the Agricultural Production Systems sIMulator (APSIM) model and multiple climate indices (e.g., climate suitability indices and extreme climate indices) were incorporated into SRM to predict the wheat yield in the North China Plain (NCP). The results showed that the prediction model based on the random forest (RF) algorithm outperformed the prediction models using other regression algorithms. The prediction for the wheat yield at SM (the period from the start of grain filling to the milky stage) based on RF can obtain a higher accuracy ($r = 0.86$, $RMSE = 683 \text{ kg ha}^{-1}$ and $MAE = 498 \text{ kg ha}^{-1}$). With the progression of wheat growth, the performances of yield prediction models improved gradually. The prediction of yield at FS (the period from flowering to the start of grain filling) can achieve higher precision and a longer lead time, which can be viewed as the optimum period providing the decent performance of the yield prediction and about one month's lead time. In addition, the precision of the predicted yield for the irrigated sites was higher than that for the rainfed sites. The APSIM-simulated AB had an importance of above 30% for the last three prediction events, including FIF event (the period from floral initiation to flowering), FS event (the period from flowering to the start of grain filling) and SM event (the period from the start of grain filling to the milky stage), which ranked first in the prediction model. The climate suitability indices, with a higher rank for every prediction event, played an important role in the prediction model. The winter wheat yield in the NCP was seriously affected by the low temperature events before flowering, the high temperature events after flowering and water stress. We hope that the prediction model can be used to develop adaptation strategies to mitigate the negative effects of climate change on crop productivity and provide the data support for food security.

Keywords: yield prediction; machine learning; APSIM model; climate indices; North China Plain



Citation: Zhao, Y.; Xiao, D.; Bai, H.; Tang, J.; Liu, D.L.; Qi, Y.; Shen, Y. The Prediction of Wheat Yield in the North China Plain by Coupling Crop Model with Machine Learning Algorithms. *Agriculture* **2023**, *13*, 99. <https://doi.org/10.3390/agriculture13010099>

Academic Editor: Yanqun Zhang

Received: 24 November 2022

Revised: 19 December 2022

Accepted: 26 December 2022

Published: 29 December 2022



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Food security is related to a series of major issues such as national food security, social stability and sustainable development of the national economy, which is highly concerned

by the country [1–3]. Increasing food productivity is an important measure to ensure food security. However, the trend of global warming became more severe throughout the 20th century [4,5]. Generally, climate warming can shorten the crop growth period, which negatively influences the formation of a crop yield and, ultimately, causes crop failure [6,7]. Predicting the yield can provide data support to take appropriate management for farmers. Wheat is one of the three major grain crops in China, with a wide planting range, large planting area and high yield [8]. Therefore, the studies on yield prediction in wheat are conducive to the timely and accurate grasp of the grain production status and scientific formulation of policies for the government [9,10].

The statistical regression model (SRM) directly developed statistical models based on the relationship between selected predictors and target variables to achieve the goal [11–14]. Guan et al. [15] used partial least-square regression (PLSR) to estimate the relationships between crop yield and the predictor variables. In general, the models combined with statistical regression algorithms are easy to understand and require fewer parameters, so the methods are commonly used in yield predictions worldwide [16–18]. However, with the increasing volume and dimension of observation data, it is a great challenge to fully explore the information of datasets for effective analysis and utilization. The most current SRM based on linear regression have some problems in application due to the complexity of the crop production system. For example, crop yields exhibit nonlinear responses to extreme climate events, while previous linear regression models may not perform well under frequent extreme climate conditions [19,20]. Compared with the linear regression analysis, the machine learning algorithm (MLA) is an advanced method for yield estimation that can capture nonlinear relations between the dependent and independent variables [21–24]. The MLA can explore the information of the training data, obtain a higher generalization level, and enhance the robustness and universality of the prediction model [15]. For example, Cai et al. [25] developed the prediction model for wheat in Australia by using some machine learning, while the support vector machine (SVM) algorithm performs better than other statistical regression algorithms. Hunt et al. [26] used the random forest (RF) algorithm to evaluate the crop yield and achieved a good performance. Nevertheless, MLA are not mechanistic and can not fully consider the dynamic process of crop growth.

The crop mechanism model (CM) has good mechanical properties, which is a simulation program that can dynamically describe the process of crop growth and yield formation under various environmental conditions by importing weather data, a variety of parameters, soil data and so on [19]. With the development of CM, the studies on the estimation of the crop yield increased gradually. For example, Huang et al. [27], Xiao et al. [28] and Zhang et al. [29], based on CMs, estimated the yields of maize, wheat and rice in China, respectively. However, the results of the most related studies were end-of-season yield predictions. The greatest limitation of within-season predictions is the lack of meteorological data from the prediction date to the maturity date [30]. Some studies achieved the prediction results by coupling the CM with seasonal weather forecasts. Pagani et al. [31] developed a high-resolution integrated prediction system for rice yield at the district level based on the combination of the WARM model, weather forecasts and remote sensing images. However, the real weather conditions may deviate from the weather forecast data, thus increasing the uncertainty of the prediction model [32].

We can reduce unnecessary errors by combining MLA and CM for yield prediction. Feng et al. [32] used the integration of the MLA and APSIM model to predict the yield of wheat under rainfed conditions in Southeastern Australia, and the hybrid model obtained a decent yield prediction at one month leading time before harvest. Nevertheless, there are few studies on using a hybrid model to predict crop yields under irrigated conditions. Furthermore, CMs can simulate the effects of complicated climate conditions on crop growth to a certain extent but are not sufficient. The quantity variation of key climatic factors (e.g., temperature and precipitation) can be transformed into the climate suitability of crop growth based on the membership function method in fuzzy mathematics [33,34]. Meanwhile, the extreme climate indices (ECIs) can quantify the destructive effects of extreme

climate events on crop growth [35,36]. The climate suitability and ECIs can be included in the hybrid model as predictive indicators to further explore the information reflected by the climate factors and improve the robustness of the hybrid model. However, there were few studies on using the combination of climate suitability and ECIs as predictive variables.

The North China Plain (NCP) is an important grain production base and occupies an important position in the national grain production in China [8]. In this study, we investigated the yield prediction of wheat in the NCP by using the CM and SRM. The main objectives of the study were (1) to develop the yield prediction model of wheat based on the combination of the multiple growth period-specific variables and SRM, (2) to identify the optimal lead time before maturity of yield prediction with acceptable accuracy, and (3) to evaluate the relative importance of input variables during different growth stages in the yield prediction model.

2. Data and Methods

2.1. Study Area

The NCP is delimited in the east by the sea, the west by the Taihang Mountains, the south by the main stream of the Huaihe River, and the north by the Yan Mountains (Figure 1) [37]. The region has a warm temperate monsoon climate with plenty of light and heat resources [37]. The annual precipitation is not evenly distributed, with over 70% of precipitation appearing in July through September. The main soil type in the NCP is the loam of Aeolian origin, a soil type deposited by rivers over geological periods. The NCP is an important grain production region in China, where the main cropping system is the double cropping systems of winter wheat–summer maize [38]. Winter wheat is usually planted in early or middle October and harvested in early June. We selected 20 agro-meteorological sites distributed across the NCP (Figure 1). Table 1 presents basic information for the 20 study sites, including location, irrigation condition, and wheat phenology and yield.

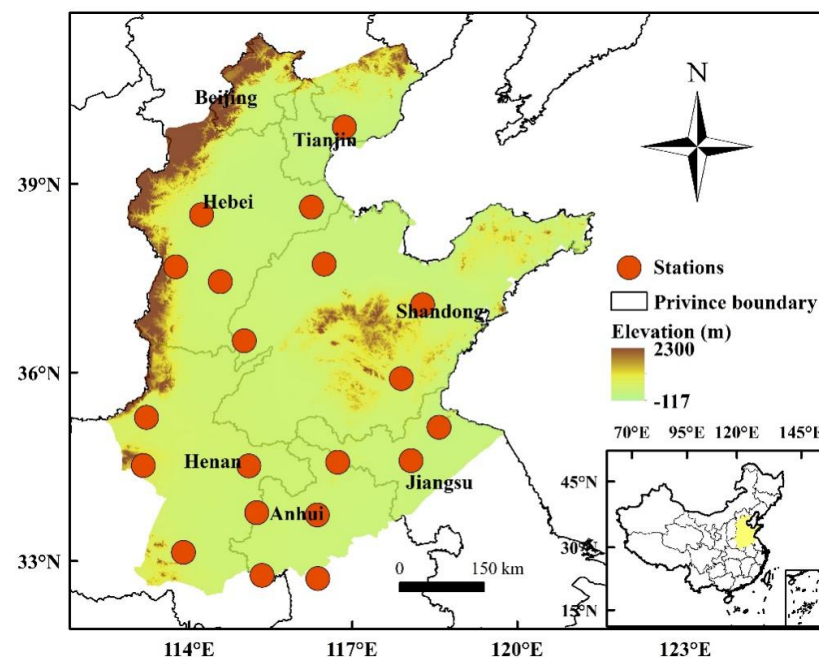


Figure 1. The spatial distribution of 20 agro-meteorological sites across the North China Plain.

Table 1. Related information about the 20 investigated sites in the study.

Site	Longitude (°E)	Latitude (°N)	Harvest Years	Irrigation	FDm (DOY)	MDm (DOY)	WYm (kg/ha)
Bozhou	115.8	33.9	2000–2010	no	114	150	5213
Dingzhou	115.0	38.3	2000–2010	yes	129	162	5728
Fuyang	115.8	32.9	2000–2010	no	112	150	6233
Ganyu	119.1	34.5	2000–2010	yes	123	159	6559
Huanghua	117.2	38.2	2004–2010	no	130	157	2924
Huimin	117.4	37.3	2000–2010	yes	128	160	6591
Juxian	118.8	35.6	2000–2010	yes	127	162	7120
Liaocheng	116.0	36.4	2000–2010	yes	124	158	5908
Luancheng	114.6	37.9	2000–2010	yes	127	162	6845
Nangong	115.3	37.3	2000–2010	yes	124	157	5580
Shangqiu	115.7	34.5	2000–2010	yes	115	150	5099
Shouxian	116.8	32.6	2000–2010	no	112	146	5316
Shuyang	118.8	34.1	2000–2010	no	123	158	6069
Suxian	116.6	33.4	2000–2010	no	116	151	6173
Tangshan	118.1	39.4	2000–2010	yes	132	166	6123
Weifang	119.2	36.8	2000–2010	yes	127	158	6017
Xinxiang	114.0	35.3	2000–2010	yes	119	151	6016
Xuzhou	117.4	34.3	2000–2010	no	118	154	7406
Zhengzhou	113.4	34.4	2000–2010	yes	112	148	5033
Zhumadian	114.1	33.0	2000–2010	no	107	143	5667

Notes: FDm, MDm and WYm denote the mean flowering date, the mean maturity date, and the mean yield for wheat during the investigated period, respectively. DOY is day of year.

2.2. Climate, Soil and Crop Data

The historical records about daily climate data, including mean temperature (Tmean), maximum temperature (Tmax), minimum temperature (Tmin), precipitation (Prec), and sunshine hours (Sh) during 2000 to 2010 for 20 agro-meteorological sites across the NCP, were obtained from China's Meteorological Administration (CMA). Soil profile data of all the sites were obtained from the 1:1 million scale soil map of China included in the Harmonized World Soil Database (HWSD) version 1.2 [39]. The climate and soil data were used to run the APSIM model.

Detailed field experimental data records included the phenology (sowing date (SD), flowering date (FD), and maturity date (MD)); grain yield; and management data at the agro-meteorological experiment sites for 2000–2010 were also obtained from CMA. The phenology data was observed by experimenters in the specific fields at the agro-meteorological experiment sites, while the grain yield was the weight of the harvested crop in the specific fields. We used the experimental crop data to calibrate and validate the crop parameters in the APSIM model.

2.3. Methodology

2.3.1. Agricultural Production Systems SIMulator (APSIM) Simulations

The APSIM model is a comprehensive model developed to simulate biophysical processes in agricultural production systems [40,41]. The APSIM model can provide an acceptable prediction accuracy of crop productivity under the combined influences of climate change, soil condition, and management measures [42,43]. In this study, the APSIM

model was implemented to simulate crop phenology, biomass, and grain yield during 2000–2010 at the 20 selected sites.

2.3.2. Climate Indices

In the study, we took account of four main growth periods, including the period from end of the juvenile stage to floral initiation (JF), the period from floral initiation to flowering (FIF), the period from flowering to the start of grain filling (FS), and the period from the start of grain filling to the milky stage (SM). We assessed the impacts of 10 extreme climate indices (ECIs) [44,45] and 3 for climate suitability (CS) [46] during different growth periods for wheat (Table 2). The calculation methods of the ECIs were shown in Table 2. The CS can further explore the information of the mean climate variables. We can develop the climate suitability model according to related studies [46].

Table 2. The information about the thirteen climate indices (CIs) used in the study.

Index	Name	Definition	Growth Stage
TS	Temperature suitability	The indicator of measurement when temperature is less or greater than physiological temperature requirement	JF, FIF, FS, SM
SS	Sunshine suitability	The indicator of measurement when sunshine is less or greater than physiological sunshine requirement	JF, FIF, FS, SM
PS	Precipitation suitability	The indicator of measurement when precipitation is less or greater than physiological water requirement	JF, FIF, FS, SM
HD	Hot days	The number of days with $T_{max} \geq 30$ °C	FS, SM
HCD	Consecutive hot days	The number of days with three or more continuous days of $T_{max} \geq 30$ °C	FS, SM
WD	Warm days	The number of days with $T_{max} > 22$ °C	JF, FIF, FS, SM
WCD	Consecutive warm days	The number of days with three or more continuous days of $T_{max} \geq 22$ °C	JF, FIF, FS, SM
FD	Frost days	The number of days with $T_{min} < 2$ °C	JF, FIF
FCD	Consecutive cold days	The number of days with three or more continuous days of $T_{min} < 2$ °C	JF, FIF
R10	Heavy precipitation days	The number of days when precipitation ≥ 10 mm	JF, FIF, FS, SM
CDD	Consecutive dry days	The number of days with three or more continuous days of daily precipitation < 1 mm	JF, FIF, FS, SM
CWD	Consecutive wet days	The number of days with three or more continuous days of daily precipitation ≥ 1 mm	JF, FIF, FS, SM
SDII	Simple daily intensity index	The ratio of total precipitation to the number of wet days (≥ 1 mm)	JF, FIF, FS, SM

Note: JF, FIF, FS, and SM denote the periods from end of the juvenile stage to floral initiation, from floral initiation to flowering, from flowering to the start of grain filling, and from the start of grain filling to the milky stage, respectively.

The sunshine suitability (SS) of wheat was calculated as follows [47–49]:

$$SS = \begin{cases} e^{-[(S_i - S_0)/b]^2} & S_i < S_0 \\ 1 & S_i \geq S_0 \end{cases} \quad (1)$$

where S_0 is the daily sunshine hours when the percentage of the daily sunshine hours reaches 70%, S_i is the daily sunshine hours (h), and b is a constant that can be determined according to the climatic conditions across the NCP and relevant studies [49,50]. The values for b at different growth periods are shown in Table 3. The arithmetic mean of the daily SS for a specific growth period is the SS for the corresponding period.

Table 3. Values of related parameters for calculating the sunshine suitability (SS), temperature suitability (TS), and precipitation suitability (PS) at four growth periods of wheat.

Parameters	JF	FIF	FS	SM
b	4.26	4.5	4.61	4.96
T ₁	−2	3	8	12
T ₀	5	10	16	20
T ₂	10	20	27	30
Kc	0.55	0.8	1.05	1.0

Note: JF, FIF, FS, and SM denote the periods from the end of the juvenile stage to floral initiation, from floral initiation to flowering, from flowering to the start of grain filling, and from start of the grain filling to the milky stage, respectively.

The temperature suitability (TS) of wheat was calculated as follows [47–49]:

$$TS = \frac{[(T_i - T_1)(T_2 - T_i)^B]}{[(T_0 - T_1)(T_2 - T_0)^B]} \quad (2)$$

$$\text{Among } \left\{ B = \frac{(T_2 - T_0)}{(T_0 - T_1)} \right\} \quad (3)$$

where T_i is the daily mean temperature (°C), T₀ is the optimal temperature (°C) at different growth periods, T₁ is the lower limit temperature (°C) at different growth periods, and T₂ is the upper limit temperature (°C) at different growth periods. The specific values of T₀, T₁, and T₂ refer to the climatic conditions across the NCP and relevant studies [49,51]. The values for T₀, T₁, and T₂ at different growth periods are listed in Table 3. The arithmetic mean of the daily TS for a specific growth period is the TS for the corresponding period.

The precipitation suitability (PS) of wheat was calculated as follows [52]:

$$PS = \begin{cases} P/P_0 & P_i < P_0 \\ P_0/P & P_i \geq P_0 \end{cases} \quad (4)$$

where P is precipitation (mm), and P₀ is the physiological water requirement of crops, which can be calculated as follows:

$$P_0 = Kc \bullet ET_0 \quad (5)$$

where Kc is the crop coefficient, and ET₀ is the reference crop evapotranspiration (mm). The Kc values of wheat at different growth stages listed in Table 3 are determined according to the relevant studies [53,54]. The ET₀ values of wheat are calculated based on the Penman–Monteith formula [54].

2.3.3. Regression Models

Two machine learning algorithms, i.e., random forest (RF) and light gradient boosting machine (LGB), were selected to predict the wheat yield. RF is an ensemble learning algorithm [26,55], which creates multiple decision trees in a random way and applies them in training samples. Among all the current algorithms, RF has high accuracy and stability, which can effectively process input samples with large data volumes and high-dimensional features. LGB is an implementation of the gradient boosting decision tree, which is essentially based on decision tree training integration to gain the optimal model [56,57]. The LGB model uses the histogram algorithm to find the best branching point, which greatly improves the training speed of the model. At the same time, LGB optimizes the growth strategy of the decision tree and uses the leaf-wise algorithm with depth limitation to create the decision tree, which can reduce the unnecessary amount of computation. In addition,

multiple linear regression (MLR) was selected as the benchmark model in this study to compare with the above two machine learning models.

2.3.4. The Framework for the Procedures

The diagram for the procedures in this study is shown in Figure 2. We developed a yield-predicting system based on multi-source environmental data using the APSIM model and regression models (MLR and RF). Firstly, the APSIM model was calibrated and validated based on observed phenology data and grain yield data at the selected sites. Then, we ran the implemented model to obtain the biomass and main growth stages, including the end of the juvenile stage, floral initiation, flowering, start of grain filling, and the milky stage. The main growth stages were used to calculate the 13 CIs. We aggregated the APSIM-accumulated biomass (AB) and climate variables into four groups by different growth periods. In the study, four prediction events (JF, FIF, FS, and SM) were triggered successively, while the predictive indicators were added with crop growth progression. Therefore, the number of predictive indicators would increase with progressing phases from JF to SM. Furthermore, we conducted “leave-one-year-out” experiments [25,58] for 2000–2010 to test the performances of the yield prediction models. Finally, the importance values for the input characteristic variables were analyzed based on the RF model and LGB model.

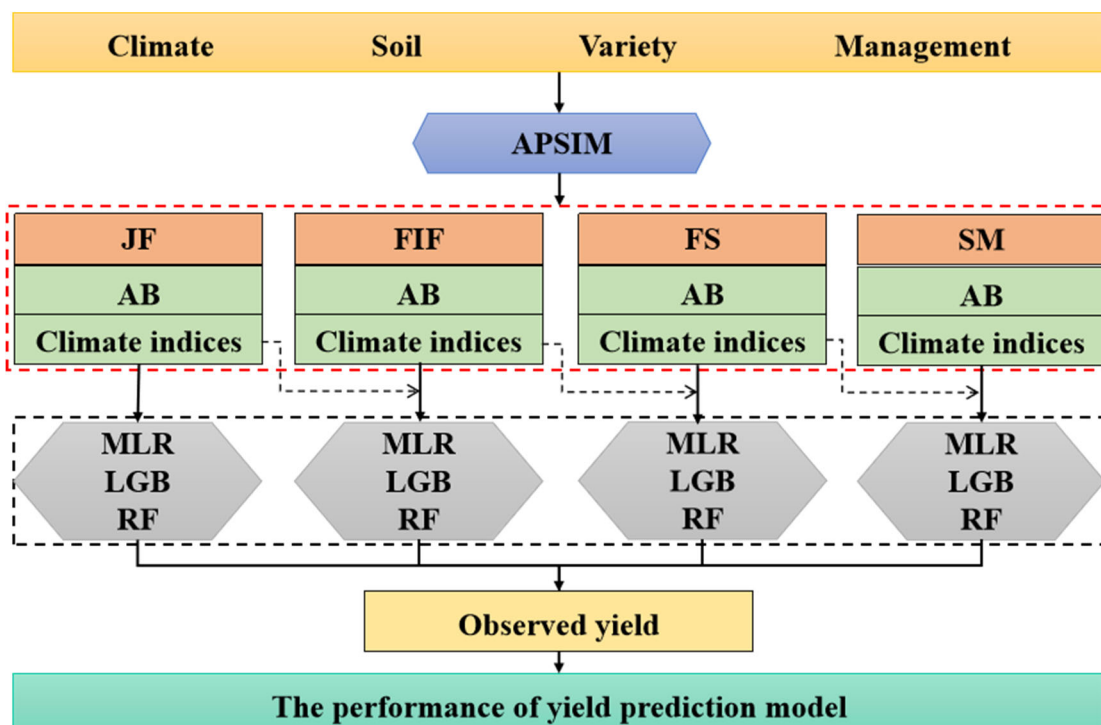


Figure 2. The diagram for the procedures used in this study, where JF, FIF, FS, and SM were the periods from the end of the juvenile stage to floral initiation, from floral initiation to flowering, from flowering to the start of grain filling, and from the start of grain filling to the milky stage, respectively. AB was accumulated biomass; MLR, LGB, and RF the multiple linear regression, light gradient boosting machine, and random forest, respectively.

2.3.5. Model Performance Assessment

The performance of the yield prediction model was validated by calculating the root mean square error (RMSE), Pearson’s correlation coefficient (r), and mean absolute error

(MAE) between the estimated data and the observed data. The calculation formulas were as follows:

$$r = \frac{\sum_{i=1}^n (O_i - \bar{O})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}} \quad (6)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - S_i)^2} \quad (7)$$

$$\text{MAE} = \frac{\sum_{i=1}^n |O_i - S_i|}{n} \quad (8)$$

where O_i , S_i , \bar{O} , \bar{S} , and n represent the observed data, estimated data, mean value of the observed data, mean value of the estimated data, and the number of samples, respectively.

3. Results

3.1. Validation of the APSIM Model

The comparison of the observed and APSIM-simulated values of the flowering date (FD), maturity date (MD), and yield from 2000 to 2010 at the 20 sites is shown in Figure 3. The simulated FD and MD were in good agreement with observed values. The r values for the simulated and observed values of FD and MD were 0.78 and 0.82, respectively. The RMSE values between the simulated and observed values of FD and MD were 5.46 d and 4.94 d, respectively. On the other hand, the simulated grain yield was consistent with the observed yield, with r of 0.81 and RMSE of 792 kg ha⁻¹. Overall, the APSIM model can provide an acceptable assessment for the phenology and grain yield of wheat. Therefore, the simulation results from the APSIM model for wheat phenology and grain yield were reliable, and we could use the simulations to develop a hybrid model for predicting the wheat yield.

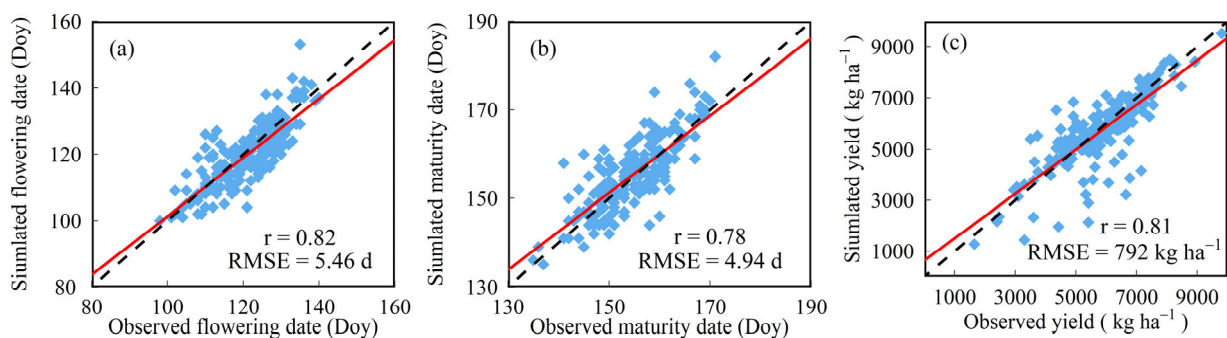


Figure 3. Comparison of observed and APSIM-simulated values of the flowering date (a), maturity date (b), and yield (c) from 2000 to 2010 at the 20 sites across the North China Plain. Red lines are the linear regression fit. Dashed lines represent the 1:1 lines.

3.2. The Model Performance and Optimum Leading Time for Yield Prediction

We developed a hybrid model to predict wheat yield based on the APSIM-simulated AB, climate indices at different growth stages and regression algorithms. The performances of three regression models are shown in Figures 4 and 5. In the early stage, the yield prediction accuracy of the three regression models was generally lower, with RMSE values of above 1000 kg ha⁻¹ and MAE values of more than 700 kg ha⁻¹ (Figures 4a,e,i and 5a,c,e). With the progression of wheat growth period, the input variables also increased, and the performances of the prediction models improved further. From JF to SM, the prediction accuracy increased significantly for the three regression models. For the MLR model, r increased from 0.22 to 0.79, RMSE decreased from 1237 kg ha⁻¹ to 778 kg ha⁻¹, and MAE decreased from 957 kg ha⁻¹ to 619 kg ha⁻¹ (Figures 4a–d and 5b). Compared with the machine learning model, MLR was less effective in predicting the wheat yield. The machine

learning model can capture the nonlinear relationship between the characteristic variables and the yield, and the overall performance of the machine learning models was good, especially the RF model. For the RF model, r increased from 0.66 to 0.86, RMSE decreased from 1026 kg ha⁻¹ to 683 kg ha⁻¹, and MAE decreased from 756 kg ha⁻¹ to 498 kg ha⁻¹ (Figures 4i–l and 5f). The performances of the yield prediction models improved gradually with the development of crop growth. However, the tradeoff between the accuracy and leading time needs to be taken into account. The yield prediction at JF will achieve the target of the prediction approximate with three months leading time before the maturity but with a poor performance ($r < 0.66$). The yield prediction at SM outperformed the yield prediction at other growth periods, while the leading time of the prediction decreased to below 15 d. A higher precision and longer lead time were taken into consideration for the prediction of the yield at FS. Therefore, FS can be regarded as the optimal period, providing the best performance of yield prediction and about one month of lead time.

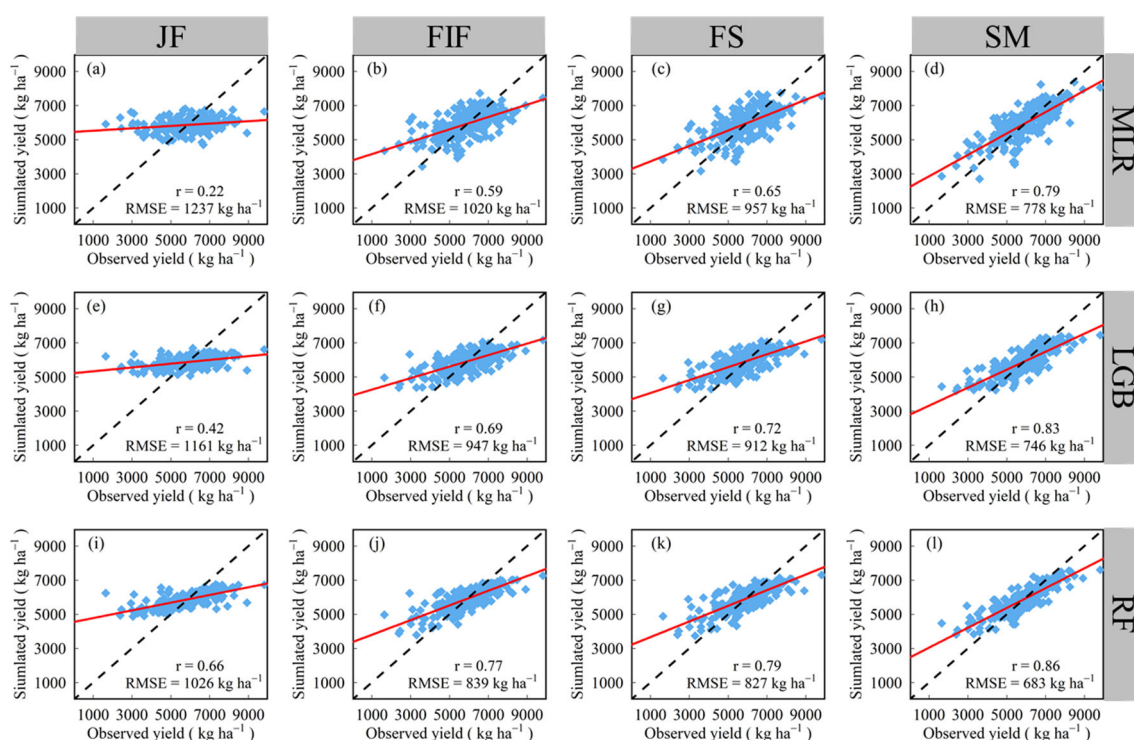


Figure 4. Comparison of the observed and predicted wheat yields for the period from the end of the juvenile stage to floral initiation (JF) (a,e,i), from floral initiation to flowering (FIF) (b,f,j), from the start of grain filling to the milky stage (FS) (c,g,k), and from flowering to the start of grain filling (SM) (d,h,l) from multiple linear regression (MLR) (a–d), light gradient boosting machine (LGB) (e–h), and random forest (RF) (i–l). Red lines are the linear regression fit. Dashed lines represent the 1:1 lines.

We compared the performance of the predicted yield across the study sites under irrigated conditions with the performance of the predicted yield across the study sites under rainfed conditions (Figures 6 and 7). The errors of the predicted yield from three regression models at all growth periods for the study sites under irrigated conditions (MAE ranged from 419 kg ha⁻¹ to 789 kg ha⁻¹) were lower than those for the rainfed sites (MAE ranged from 624 kg ha⁻¹ to 1130 kg ha⁻¹) (Figures 6 and 7). The accuracy of the predicted yield for the irrigated sites was higher than that for the rainfed sites. The water shortage caused by drought limited photosynthesis and carbon allocation, which was not conducive to the formation of the crop yield and affected the prediction accuracy [59,60]. However, the impacts of water stress on the crop yield was reduced by irrigation, which improved the accuracy of the yield prediction under irrigated conditions [61–64]. Nevertheless, the pre-

dicted yield for the study sites under irrigated conditions were underestimated compared to the observed yield, while the predicted yield for the rainfed sites were overestimated compared to the observed yield (Figures 6a,c,e and 7a,c,e).

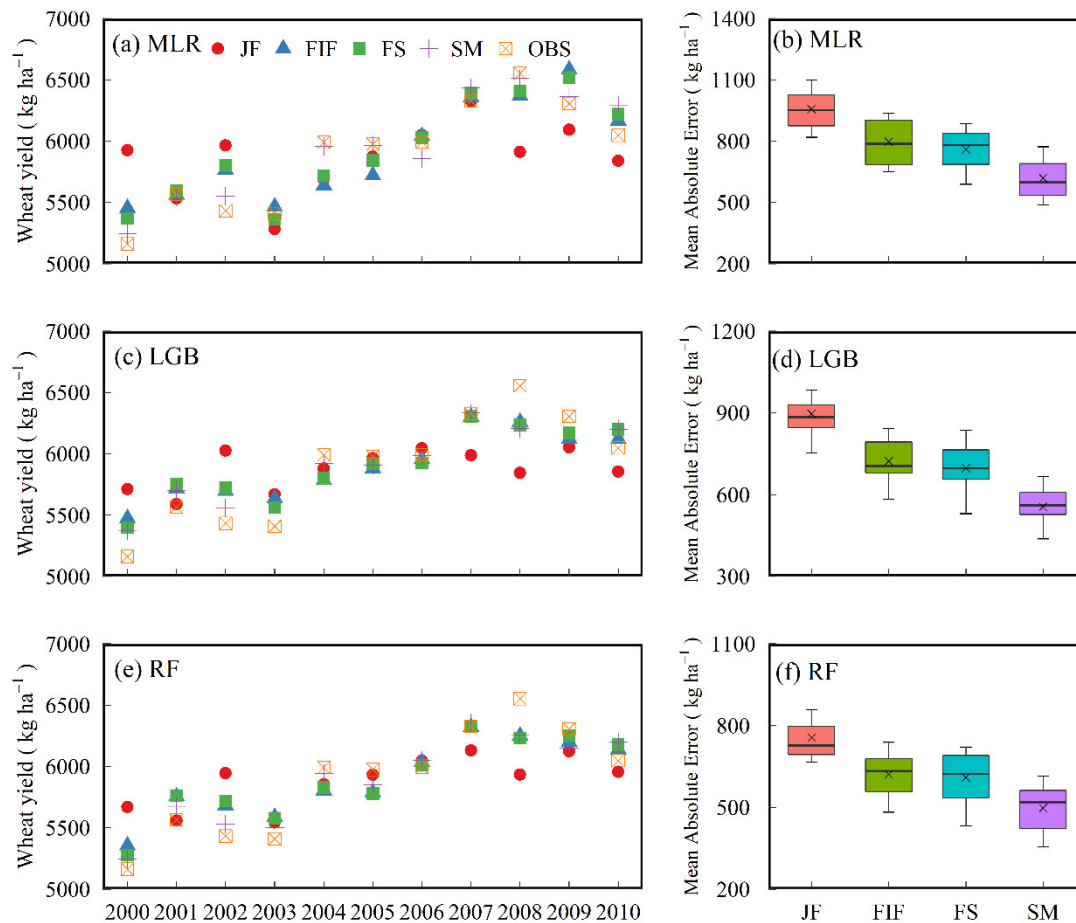


Figure 5. Time series of observed and predicted wheat yields across the 20 investigated sites based on the four prediction events from multiple linear regression (MLR) (a,b), light gradient boosting machine (LGB) (c,d), and random forest (RF) (e,f). Wheat yields for each year were averaged across the 20 investigated sites. Data were generated from the “leave-one-year-out” cross-validation procedure from the three regression models. JF, FIF, FS, and SM were the periods from the end of the juvenile stage to floral initiation, from floral initiation to flowering, from flowering to the start of grain filling, and from the start of grain filling to the milky stage, respectively.

3.3. Relative Importance of Selected Predictors at Different Growth Stages

The RF model and LGB model were used to assess the importance of the input characteristic variables in the yield prediction model. The relative importance of the input predictors as determined from the average of the LGB model and RF model for each prediction event is shown in Figure 8. With the crop growth and progression, the importance of the APSIM-simulated accumulated biomass (AB) increased rapidly, while the importance of AB at the last three prediction events was over 30% (Figure 8). For the CIs, the climate suitability indices were most important for the yield prediction at the early prediction event, such as TS and SS (Figure 8a). The roles of the climate suitability indices in the prediction model should not be ignored, though some extreme climate indices had higher importance than the climate suitability indices in the last three prediction events (Figure 8b–d). In the middle prediction events (FIF and FS), SDII and FD at FIF generally ranked high in the climate indices, which may be because the wheat yield was very sensitive to low-temperature stress and water stress before flowering (Figure 8b,c). However, SDII and HCD at SM ranked first in the late prediction events, suggesting that the impact of

heat stress and water stress after flowering on the wheat yield was more significant than low-temperature stress and water stress before flowering (Figure 8b–d).

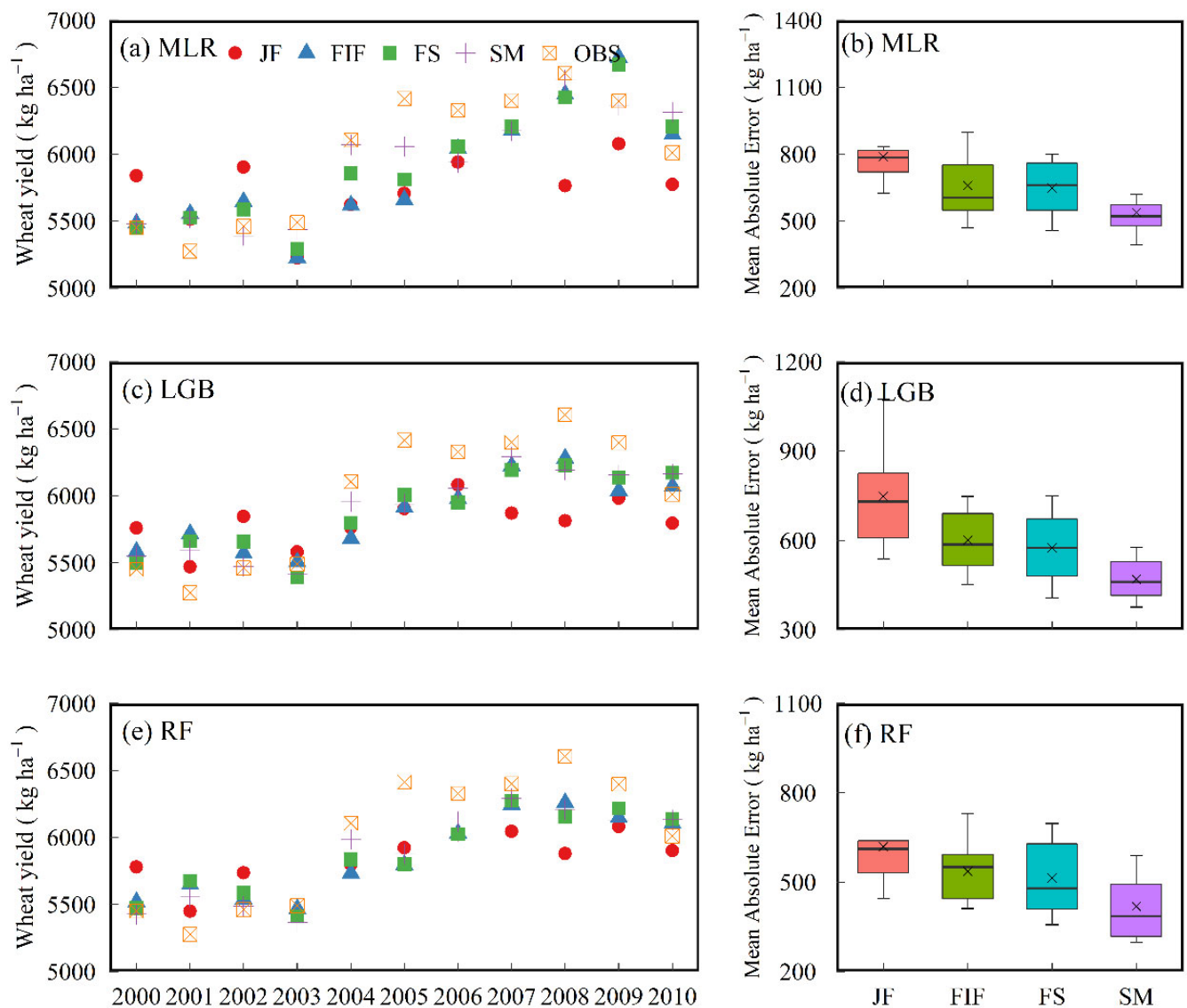


Figure 6. Time series of observed and predicted wheat yields across the investigated sites under irrigated conditions based on the four prediction events from multiple linear regression (MLR) (a,b), light gradient boosting machine (LGB) (c,d), and random forest (RF) (e,f). Wheat yields for each year were averaged across the investigated sites under irrigated conditions. Data were generated from the “leave-one-year-out” cross-validation procedure from the three regression models. JF, FIF, FS, and SM were the periods from the end of the juvenile stage to floral initiation, from floral initiation to flowering, from flowering to the start of grain filling, and from the start of grain filling to the milky stage, respectively.

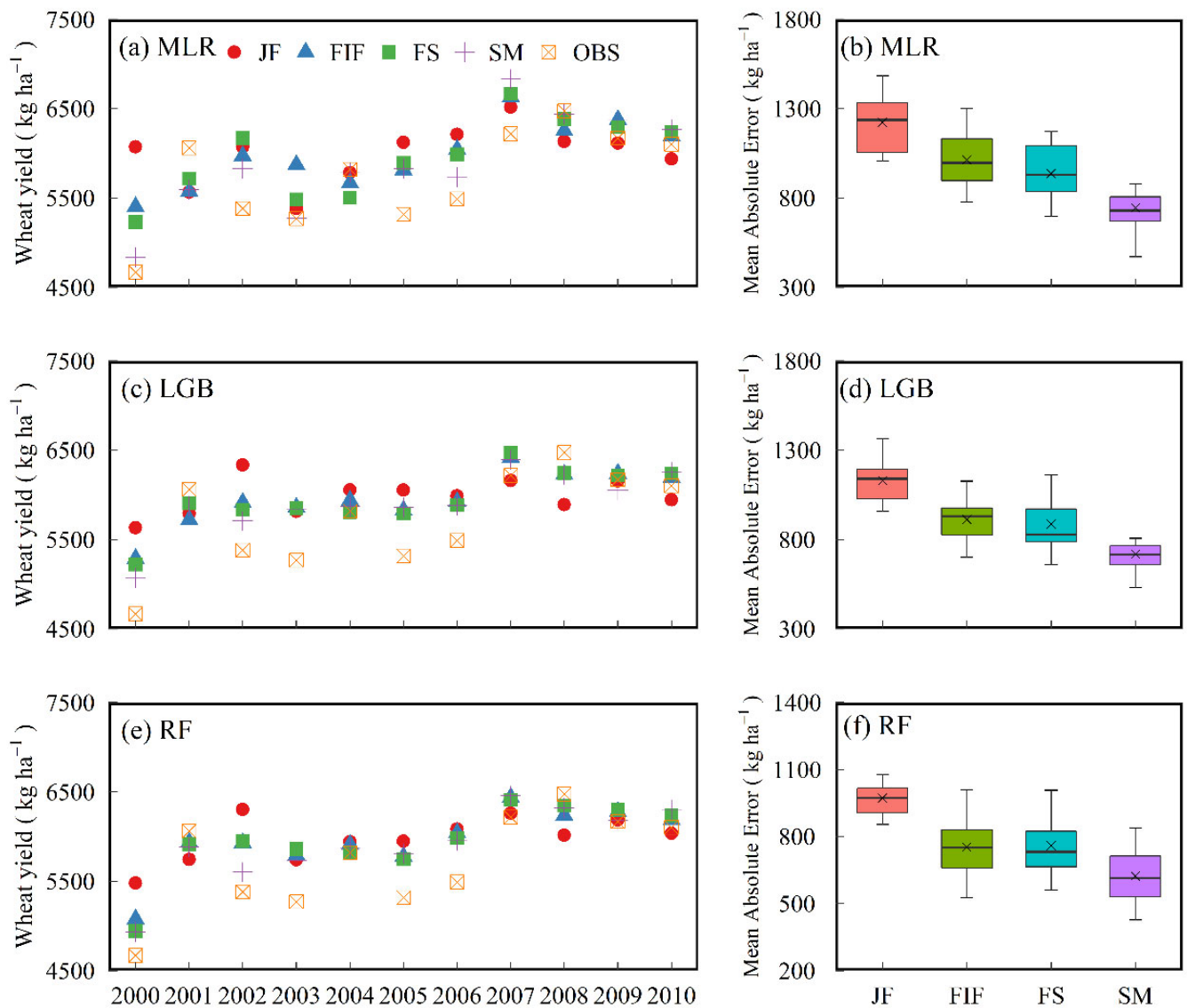


Figure 7. Time series of observed and predicted wheat yields across the investigated sites under rainfed conditions based on the four prediction events from multiple linear regression (MLR) (a,b), light gradient boosting machine (LGB) (c,d), and random forest (RF) (e,f). Wheat yields for each year were averaged across the investigated sites under rainfed conditions. Data were generated from the “leave-one-year-out” cross-validation procedure from the three regression models. JF, FIF, FS, and SM were the periods from the end of the juvenile stage to floral initiation, from floral initiation to flowering, from flowering to the start of grain filling, and from the start of grain filling to the milky stage, respectively.

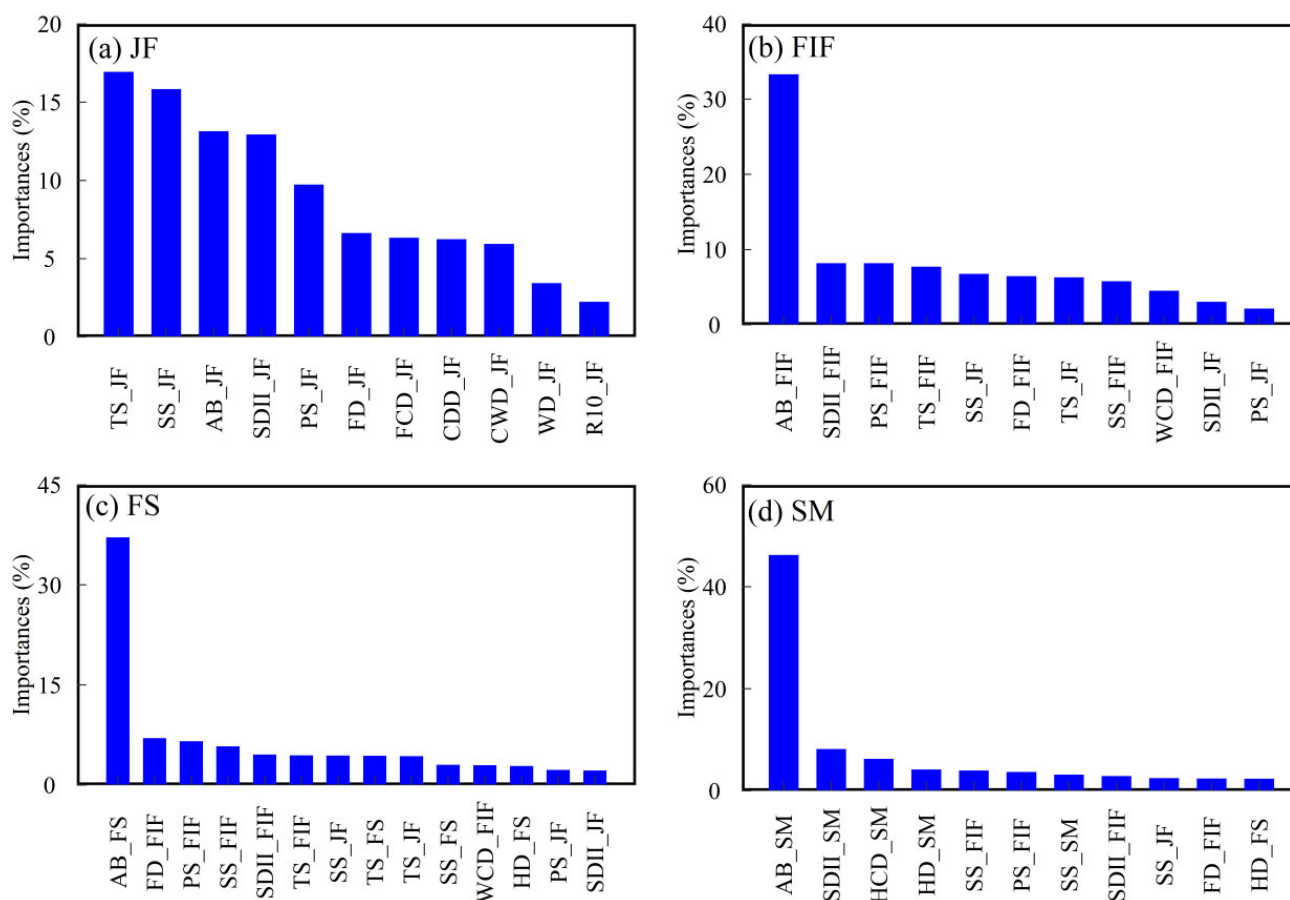


Figure 8. Relative importance of the input predictors as determined from the average of LGB (light gradient boosting machine) model and RF (random forest) model for the period from the end of the juvenile stage to floral initiation (JF) (a), from floral initiation to flowering (FIF) (b), from flowering to the start of grain filling (FS) (c), and from the start of grain filling to the milky stage (SM) (d). The results are normalized to sum 100% and shown in decreasing order in the figure (The input predictors lower than 2% were not shown in the figure).

4. Discussion

A crop model can dynamically describe the process of crop growth and development under various environmental conditions [65]. A growing body of studies have investigated the effects of climate change during the past few decades on crop phenology and yield using various crop models to develop adaptive measures (such as adjustment of the sowing date and renewal of crop variety) for reducing the yield loss [66–68]. However, there were fewer studies on the yield prediction using the crop model due to the limitation of the meteorological data. Some studies used the combination of statistical regression models and crop models to estimate the crop yield. For example, Everingham et al. [69] built one prediction model for sugarcane yield by incorporating the biomass simulated by the crop model and several climate indices into the RF algorithm and obtained a high accuracy. Similarly, Feng et al. [32] conducted the study on the yield prediction for wheat in South-Eastern Australia through combining the APSIM model and RF model, obtaining a high accuracy ($r = 0.87$, $RMSE = 640 \text{ kg ha}^{-1}$). In this study, we developed a hybrid model for the yield prediction of the wheat coupling crop model and several statistical regression models. The yield prediction model based on the crop model and RF algorithm outperformed the yield prediction model based on the crop model and other regression algorithms (MLR and LGB), with r of 0.86 and $RMSE$ of 683 kg ha^{-1} . The precision of the study was similar to the related study [32]. This may be because that RF algorithm has a

strong ability of data processing, which improved the accuracy and robustness of the yield prediction based on the RF algorithm [70–72].

Global climate change has a significant impact on social economy and the natural environment, especially on agricultural production [73–80]. Different crops have different demands for climate resources, and more or less, climate resources are not conducive to crop growth and development [81,82]. As compared with extreme climate events, mean climate conditions generally made more contributions to the variations of wheat growth in the NCP [83]. The climate suitability can be used to estimate the sensitivity of crops to climate factors, such as mean temperature, precipitation, and sunshine, and there is a certain correlation between climate suitability and climatic yield [84,85]. Climate suitability can further explore the information reflected by the mean climate conditions, though the crop model can simulate the effects of the mean climate conditions on crop growth to some extent. In this study, the climate suitability indices (TS, SS, and PS) played an important role in predicting the final yield of wheat, which generally ranked high in the models for every prediction event (Figure 8). The roles of climate suitability indices in the prediction model should not be ignored.

Extreme temperature events have a negative influence on crop growth and yield formation, which could cause crop yield loss [2,86–88]. The low-temperature events before flowering and high-temperature events after flowering are two major extreme temperature events affecting winter wheat [89,90]. Xiao et al. [91] found that there was the greatest frost duration and intensity in the NCP, which suffered the largest yield losses due to spring frost events. Warm temperatures can improve the growth of crops before the temperature reaches the threshold, but yields will abruptly diminish subsequently [92,93]. Around flowering or the grain filling period, extreme high temperature could affect pollination, reduce male fertility, and the efficiency of grain yield, and a large yield loss would be caused by continuous heat stress [94–97]. Bai et al. [98] found that heat stress after flowering significantly negatively impacts wheat production in the NCP, while the wheat yield might have a higher frequency of exposure to extreme high-temperature stress in the future. The findings of the study showed that FD at FIF ranked high for the middle prediction events (FIF and FS), while HCD at SM ranked first in the climate indices for the late prediction events (Figure 8). Low-temperature events before flowering and heat stress after flowering are the main natural disasters affecting wheat growth in the NCP [90,99]. It is of great significance to take appropriate measures to alleviate the negative effects of these disasters on crops.

Drought is also closely correlated to agricultural production [100–103]. In this study, the rank of SDII related to water stress was consistently high for all prediction events, indicating that water stress has a significant impact on wheat yield in the North China Plain. Water stress can affect the coupling mechanism of environmental driving factors and crop yield, while it is difficult to achieve the acceptable yield prediction in the rainfed system [61]. The predicted yield for the sites under rainfed conditions would be overestimated due to the water stress, while irrigation can effectively reduce the effect of drought on the crop yield and increase the accuracy of the crop yield prediction [62,83]. In the study, the predicted yield for the sites in the rainfed system was overestimated, while the MAE of the predicted yield for the sites under the irrigated condition was significantly lower than that of the sites in the rainfed system (Figures 6 and 7). More predictive variables may need to be incorporated into the hybrid model to improve the performance of the model under irrigated conditions.

There are still some uncertainties and limitations in our study. The RF model is more dependent on data. Sufficient data samples are conducive to improving the accuracy and robustness of the model, while the lack of training samples may lead to overfitting and increase the uncertainty of the model [104]. The data processing ability of machine learning algorithms can fully function by obtaining more yield samples, while the performance of the model can be improved further. Furthermore, the model developed in this study is limited to the yield prediction at the site scale, which is difficult to be applied in a large-scale

region. Lobell et al. [105] developed a scalable satellite-based crop yield mapper (SCYM) based on satellite images and crop models, which successfully explained 35% of the maize yield variation and 32% of the soybean yield variation in the study area. In the future, we can incorporate the SCYM model into the hybrid model to predict the crop yield at a large-scale region. This is a study direction with great development potential.

5. Conclusions

Based on the APSIM-simulated AB, climate indices at different growth stages, and statistical regression algorithms, we developed a hybrid model to predict wheat yields in the NCP. The results showed that the prediction model based on machine learning algorithms outperformed the prediction models using MLR regression, especially the RF algorithm. The performances of the yield prediction models improved gradually with the development of crop growth. A higher precision and longer lead time were taken into consideration for the prediction of the yield at FS. The FS can be regarded as the optimal period, providing the acceptable performance of yield prediction and about one month lead time. Moreover, the accuracy of the predicted yield for the irrigated sites was higher than that for the rainfed sites. The APSIM-simulated AB dominated the last three prediction events, with the importance above 30%. The climate suitability indices played an important role in predicting the wheat yield, with high rankings for every prediction event. Among extreme climate events, the low temperature events before flowering, high temperature events after flowering, and water stress were major extreme climate events affecting the winter wheat yield.

In general, the hybrid model can be used to predict the wheat yield under both rainfed and irrigated conditions in the NCP. This model is helpful in developing adaptation strategies to alleviate the negative effects of climate change on crop productivity and improve agricultural risk management. Nevertheless, the hybrid model is dependent on the quantity and quality of the data samples. Furthermore, the model developed in this study is limited in the yield prediction at the large-scale region. In the next study, we can incorporate the SCYM model into the hybrid model to predict crop yield at large-scale regions.

Author Contributions: Conceptualization, D.X.; methodology and data analysis, Y.Z.; writing—original draft preparation, Y.Z.; and writing—review and editing, D.X., H.B., J.T., D.L.L., Y.Q. and Y.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research and the APC were funded by a grant from the Hebei Provincial Science Foundation for Distinguished Young Scholars (No. D2022205010), the National Natural Science Foundation of China (No. 41901128), the High-level Talents Training and Subsidy Project of Hebei Academy of Science (2022G04), and the Technology Program of Hebei Academy of Sciences (22102).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Godfray, H.C.J.; Beddington, J.R.; Crute, I.R.; Haddad, L.; Lawrence, D.; Muir, J.F.; Pretty, J.; Robinson, S.; Thomas, S.M.; Toulmin, C. Food Security: The Challenge of Feeding 9 billion People. *Science* **2010**, *327*, 812–818. [[CrossRef](#)] [[PubMed](#)]
2. Bailey-Serres, J.; Parker, J.E.; Ainsworth, E.A.; Oldroyd, G.E.D.; Schroeder, J.I. Genetic strategies for improving crop yields. *Nature* **2019**, *575*, 109–118. [[CrossRef](#)] [[PubMed](#)]
3. Kotz, M.; Levermann, A.; Wenz, L. The effect of rainfall changes on economic production. *Nature* **2022**, *601*, 223–227. [[CrossRef](#)] [[PubMed](#)]
4. IPCC. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., et al., Eds.; Cambridge University Press: Cambridge, UK, 2021.
5. Piao, S.; Ciais, P.; Huang, Y.; Shen, Z.; Peng, S.; Li, J.; Zhou, L.; Liu, H.; Ma, Y.; Ding, Y.; et al. The impacts of climate change on water resources and agriculture in China. *Nature* **2010**, *467*, 43–51. [[CrossRef](#)] [[PubMed](#)]

6. Estrella, N.; Sparks, T.H.; Menzel, A. Trends and temperature response in the phenology of crops in Germany. *Glob. Chang. Biol.* **2007**, *13*, 1737–1747. [[CrossRef](#)]
7. Li, E.; Zhao, J.; Pullens, J.W.M.; Yang, X. The compound effects of drought and high temperature stresses will be the main constraints on maize yield in Northeast China. *Sci. Total Environ.* **2022**, *812*, 152461. [[CrossRef](#)]
8. National Bureau of Statistics of China. *China Rural Statistical Yearbook*; China Statistics Press: Beijing, China, 2020.
9. Becker-Reshef, I.; Vermote, E.; Lindeman, M.; Justice, C. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sens. Environ.* **2010**, *114*, 1312–1323. [[CrossRef](#)]
10. Tao, F.; Zhang, L.; Zhang, Z.; Chen, Y. Designing wheat cultivar adaptation to future climate change across China by coupling biophysical modelling and machine learning. *Eur. J. Agron.* **2022**, *136*, 126500. [[CrossRef](#)]
11. Bolton, D.K.; Friedl, M.A. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agric. Forest Meteorol.* **2013**, *173*, 74–84. [[CrossRef](#)]
12. Franch, B.; Vermote, E.F.; Becker-Reshef, I.; Claverie, M.; Huang, J.; Zhang, J.; Justice, C.; Sobrino, J.A. Improving the timeliness of winter wheat production forecast in the United States of America, Ukraine and China using MODIS data and NCAR Growing Degree Day information. *Remote Sens. Environ.* **2015**, *161*, 131–148. [[CrossRef](#)]
13. Erfanian, S.; Ziaullah, M.; Tahir, M.A.; Ma, D. How does justice matter in developing supply chain trust and improving information sharing—an empirical study in Pakistan. *Int. J. Manuf. Technol. Manag.* **2021**, *35*, 354–368. [[CrossRef](#)]
14. Razaq, A.; Xiao, M.; Zhou, Y.; Anwar, M.; Liu, H.; Luo, F. Towards Sustainable Water Use: Factors Influencing Farmers' Participation in the Informal Groundwater Markets in Pakistan. *Front. Environ. Sci.* **2022**, *10*, 944156. [[CrossRef](#)]
15. Guan, K.; Wu, J.; Kimball, J.S.; Anderson, M.C.; Frothingham, S.; Li, B.; Hain, C.R.; Lobell, D.B. The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields. *Remote Sens. Environ.* **2017**, *199*, 333–349. [[CrossRef](#)]
16. Mkhabela, M.S.; Bullock, P.; Raj, S.; Wang, S.; Yang, Y. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. *Agric. Forest Meteorol.* **2011**, *151*, 385–393. [[CrossRef](#)]
17. Michel, L.; Makowski, D. Comparison of statistical models for analyzing wheat yield time series. *PLoS ONE* **2013**, *8*, e78615. [[CrossRef](#)]
18. Satir, O.; Berberoglu, S. Crop yield prediction under soil salinity using satellite derived vegetation indices. *Field Crops Res.* **2016**, *192*, 134–143. [[CrossRef](#)]
19. Feng, P.; Wang, B.; Liu, D.L.; Waters, C.; Yu, Q. Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia. *Agric. Forest Meteorol.* **2019**, *275*, 100–113. [[CrossRef](#)]
20. Li, Y.; Guan, K.; Yu, A.; Peng, B.; Zhao, L.; Li, B.; Peng, J. Toward building a transparent statistical model for improving crop yield prediction: Modeling rainfed corn in the U.S. *Field Crops Res.* **2019**, *234*, 55–65. [[CrossRef](#)]
21. Kamir, E.; Waldner, F.; Hochman, Z. Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *ISPRS J. Photogramm.* **2020**, *160*, 124–135. [[CrossRef](#)]
22. Oliveira, R.A.; Näsi, R.; Niemeläinen, O.; Nyholm, L.; Honkavaara, E. Machine learning estimators for the quantity and quality of grass swards used for silage production using drone-based imaging spectrometry and photogrammetry. *Remote Sens. Environ.* **2020**, *246*, 111830. [[CrossRef](#)]
23. Wan, L.; Cen, H.; Zhu, J.; Zhang, J.; Zhu, Y.; Sun, D.; Du, X.; Zhai, L.; Weng, H.; Li, Y.; et al. Grain yield prediction of rice using multi-temporal UAV-based RGB and multispectral images and model transfer—A case study of small farmlands in the South of China. *Agric. Forest Meteorol.* **2020**, *291*, 108096. [[CrossRef](#)]
24. Erfanian, S.; Zhou, Y.; Razaq, A.; Abbas, A.; Safeer, A.A.; Li, T. Predicting Bitcoin (BTC) Price in the Context of Economic Theories: A Machine Learning Approach. *Entropy* **2022**, *24*, 1487. [[CrossRef](#)]
25. Cai, Y.; Guan, K.; Lobell, D.; Potgieter, A.B.; Wang, S.; Peng, J.; Xu, T.; Asseng, S.; Zhang, Y.; You, L.; et al. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agric. Forest Meteorol.* **2019**, *274*, 144–159. [[CrossRef](#)]
26. Hunt, M.L.; Blackburn, G.A.; Carrasco, L.; Redhead, J.W.; Rowland, C.S. High resolution wheat yield mapping using Sentinel-2. *Remote Sens. Environ.* **2019**, *233*, 111410. [[CrossRef](#)]
27. Huang, S.; Lv, L.; Zhu, J.; Li, Y.; Tao, H.; Wang, P. Extending growing period is limited to offsetting negative effects of climate changes on maize yield in the North China Plain. *Field Crops Res.* **2018**, *215*, 66–73. [[CrossRef](#)]
28. Xiao, D.; Shen, Y.; Zhang, H.; Moiwo, J.P.; Qi, Y.; Wang, R.; Pei, H.; Zhang, Y.; Shen, H. Comparison of winter wheat yield sensitivity to climate variables under irrigated and rain-fed conditions. *Front. Earth Sci.* **2015**, *10*, 444–454. [[CrossRef](#)]
29. Zhang, T.; Huang, Y.; Yang, X. Climate warming over the past three decades has shortened rice growth duration in China and cultivar shifts have further accelerated the process for late rice. *Glob. Chang. Biol.* **2013**, *19*, 563–570. [[CrossRef](#)]
30. Basso, B.; Liu, L. Chapter Four—Seasonal crop yield forecast: Methods, applications, and accuracies. In *Advances in Agronomy*; Sparks, D.L., Ed.; Academic Press: Cambridge, MA, USA, 2019; Volume 154, pp. 201–255.
31. Pagani, V.; Guarneri, T.; Busetto, L.; Ranghetti, L.; Boschetti, M.; Movedi, E.; Campos-Taberner, M.; Garcia-Haro, F.J.; Katsantonis, D.; Stavrakoudis, D.; et al. A high-resolution, integrated system for rice yield forecasting at district level. *Agric. Syst.* **2019**, *168*, 181–190. [[CrossRef](#)]
32. Feng, P.; Wang, B.; Liu, D.L.; Waters, C.; Xiao, D.; Shi, L.; Yu, Q. Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique. *Agric. For. Meteorol.* **2020**, *285*, 107922. [[CrossRef](#)]

33. Ma, S.Q. The climatic and ecological suitability of central Jilin Province for developing maize zones. *J. Ecol.* **1990**, *54*, 40–45.
34. Gong, L.J.; Yang-Hui, J.I.; Wang, P.; Zhu, H.X.; Wang, L.L.; Wang, Q.J. Variation of Climate Suitability of Maize in Northeast of China. *J. Maize Sci.* **2013**, *21*, 140–146.
35. Lobell, D.B.; Hammer, G.L.; McLean, G.; Messina, C.; Roberts, M.J.; Schlenker, W. The critical role of extreme heat for maize production in the United States. *Nat. Clim. Chang.* **2013**, *3*, 497–501. [[CrossRef](#)]
36. Wu, J.; Han, Z.; Xu, Y.; Zhou, B.; Gao, X. Changes in Extreme Climate Events in China under 1.5–4 °C Global Warming Targets: Projections Using an Ensemble of Regional Climate Model Simulations. *J. Geophys. Res. Atmos.* **2020**, *125*, e2019JD031057. [[CrossRef](#)]
37. Xiao, D.; Liu, D.L.; Wang, B.; Feng, P.; Waters, C. Designing high-yielding maize ideotypes to adapt changing climate in the North China Plain. *Agric. Syst.* **2020**, *181*, 102805. [[CrossRef](#)]
38. Xiao, D.; Tao, F. Contributions of cultivar shift, management practice and climate change to maize yield in North China Plain in 1981–2009. *Int. J. Biometeorol.* **2016**, *60*, 1111–1122. [[CrossRef](#)]
39. Shi, X.; Yu, D.; Warner, E.; Pan, X.; Petersen, G.; Gong, Z.; Weindorf, D. Soil database of 1:1,000,000 digital soil survey and reference system of the Chinese genetic soil classification system. *Soil Surv. Horiz.* **2004**, *45*, 129–136. [[CrossRef](#)]
40. Asseng, S.; Keating, B.A.; Fillery, I.R.P.; Gregory, P.J.; Abrecht, D.G. Performance of the APSIM-wheat model in Western Australia. *Field Crops Res.* **1998**, *57*, 163–179. [[CrossRef](#)]
41. Asseng, S.; Van Keulen, H.; Stol, W. Performance and application of the APSIM Nwheat model in the Netherlands. *Eur. J. Agron.* **2000**, *12*, 37–54. [[CrossRef](#)]
42. Keating, B.A.; Carberry, P.S.; Hammer, G.L.; Probert, M.E.; Robertson, M.J.; Holzworth, D.; Huth, N.I.; Hargreaves, J.N.G.; Meinke, H.; Hochman, Z. An overview of APSIM, a model designed for farming systems simulation. *Eur. J. Agron.* **2003**, *18*, 267–288. [[CrossRef](#)]
43. Arshad, A.; Raza, M.A.; Zhang, Y.; Zhang, L.; Wang, X.; Ahmed, M.; Habib-ur-Rehman, M. Impact of Climate Warming on Cotton Growth and Yields in China and Pakistan: A Regional Perspective. *Agriculture* **2021**, *11*, 97. [[CrossRef](#)]
44. Bai, H.; Xiao, D.; Wang, B.; Liu, D.L.; Feng, P.; Tang, J. Multi-model ensemble of CMIP6 projections for future extreme climate stress on wheat in the North China plain. *Int. J. Climatol.* **2020**, *41*, E171–E186. [[CrossRef](#)]
45. Xiao, D.; Bai, H.; Liu, D.L.; Tang, J.; Wang, B.; Shen, Y.; Cao, J.; Feng, P. Projecting future changes in extreme climate for maize production in the North China Plain and the role of adjusting the sowing date. *Mitig. Adapt. Strateg. Glob. Chang.* **2022**, *27*, 1–21. [[CrossRef](#)]
46. Zhao, Y.; Xiao, D.; Bai, H.; Tang, J.; Liu, D. Future Projection for Climate Suitability of Summer Maize in the North China Plain. *Agriculture* **2022**, *12*, 348. [[CrossRef](#)]
47. Zhao, F.; Qian, H.S.; Jiao, S.X. The climatic suitability model of crop: A case study of winter wheat in Henan province. *Resour. Sci.* **2003**, *25*, 77–82.
48. Zhao, J.; Guo, J.; Xu, Y.; Mu, J. Effects of climate change on cultivation patterns of spring maize and its climatic suitability in Northeast China. *Agric. Ecosyst. Environ.* **2015**, *202*, 178–187. [[CrossRef](#)]
49. Zhao, Y.; Xiao, D.; Bai, H.; Tang, J.Z. Climatic suitability degrees of winter wheat and summer maize in the North China Plain. *Chin. J. Ecol.* **2020**, *39*, 1–12.
50. Huang, H. A study on the climatic ecology adaptability of the crop production in the red and yellow soils region of China. *J. Nat. Resour.* **1996**, *11*, 340–346.
51. Pu, J.Y.; Yao, X.Y.; Yao, R.X. Variations of summer and autumn grain crops' climatic suitability in the areas east of Yellow River in Gansu in recent 40 years. *Agric. Res. Arid. Areas* **2011**, *29*, 253–258.
52. Hou, Y.Y.; Zhang, Y.H.; Wang, L.Y.; Hou-Quan, L.; Song, Y.B. Climatic suitability model for spring maize in Northeast China. *Chin. J. Appl. Ecol.* **2013**, *24*, 3207–3212.
53. Wang, E.; Han, X. The productivity evaluation and its application of winter wheat and summer maize in Huang-Huai-Hai region. *Chin. J. Agrometeorol.* **1990**, *11*, 41–46.
54. Allen, R.; Pereira, L.; Dirck, R.; Smith, M. *Crop Evapotranspiration Guidelines for Computing Crop Water Requirements—FAO Irrigation and Drainage Paper No 56*; FAO: Rome, Italy, 1998; Volume 300, pp. 53–62.
55. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
56. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154.
57. Sun, X.; Liu, M.; Sima, Z. A novel cryptocurrency price trend forecasting model based on LightGBM. *Financ. Res. Lett.* **2020**, *32*, 101084. [[CrossRef](#)]
58. Cao, J.; Zhang, Z.; Luo, Y.; Zhang, L.; Zhang, J.; Li, Z.; Tao, F. Wheat yield predictions at a county and field scale with deep learning, machine learning, and google earth engine. *Eur. J. Agron.* **2021**, *123*, 126204. [[CrossRef](#)]
59. Chaves, M.M.; Pereira, J.S.; Maroco, J.; Rodrigues, M.L.; Ricardo, C.P.; Osorio, M.L.; Carvalho, I.; Faria, T.; Pinheiro, C. How plants cope with water stress in the field. Photosynthesis and growth. *Ann. Bot.* **2002**, *89*, 907–916. [[CrossRef](#)]
60. Sun, H.; Zhang, X.; Liu, X.; Liu, X.; Shao, L.; Chen, S.; Wang, J.; Dong, X. Impact of different cropping systems and irrigation schedules on evapotranspiration, grain yield and groundwater level in the North China Plain. *Agric. Water Manag.* **2019**, *211*, 202–209. [[CrossRef](#)]

61. Grassini, P.; Yang, H.; Cassman, K.G. Limits to maize productivity in Western Corn-Belt: A simulation analysis for fully irrigated and rainfed conditions. *Agric. Forest Meteorol.* **2009**, *149*, 1254–1265. [[CrossRef](#)]
62. Sibley, A.M.; Grassini, P.; Thomas, N.E.; Cassman, K.G.; Lobell, D.B. Testing Remote Sensing Approaches for Assessing Yield Variability among Maize Fields. *Agron. J.* **2014**, *106*, 24–32. [[CrossRef](#)]
63. Razzaq, A.; Qing, P.; Naseer, M.; Abid, M.; Anwar, M.; Javed, I. Can the informal groundwater markets improve water use efficiency and equity? Evidence from a semi-arid region of Pakistan. *Sci. Total Environ.* **2019**, *666*, 849–857. [[CrossRef](#)]
64. Razzaq, A.; Xiao, M.; Zhou, Y.; Liu, H.; Abbas, A.; Liang, W.; Naseer, M.A.U.R. Impact of Participation in Groundwater Market on Farmland, Income, and Water Access: Evidence from Pakistan. *Water* **2022**, *14*, 1832. [[CrossRef](#)]
65. De Wit, C.T. *Photosynthesis of Leaf Canopies*; Pudoc: Wageningen, The Netherlands, 1965.
66. Xiao, D.; Shen, Y.; Qi, Y.; Moiwo, J.P.; Min, L.; Zhang, Y.; Guo, Y.; Pei, H. Impact of alternative cropping systems on groundwater use and grain yields in the North China Plain Region. *Agric. Syst.* **2017**, *153*, 109–117. [[CrossRef](#)]
67. Yan, Z.; Zhang, X.; Rashid, M.A.; Li, H.; Jing, H.; Hochman, Z. Assessment of the sustainability of different cropping systems under three irrigation strategies in the North China Plain under climate change. *Agric. Syst.* **2020**, *178*, 102745. [[CrossRef](#)]
68. Zhu, G.; Liu, Z.; Qiao, S.; Zhang, Z.; Huang, Q.; Su, Z.; Yang, X. How could observed sowing dates contribute to maize potential yield under climate change in Northeast China based on APSIM model. *Eur. J. Agron.* **2022**, *136*, 126511. [[CrossRef](#)]
69. Everingham, Y.; Sexton, J.; Skocaj, D.; Inman-Bamber, G. Accurate prediction of sugarcane yield using a random forest algorithm. *Agron. Sustain. Dev.* **2016**, *36*, 1–9. [[CrossRef](#)]
70. Wang, L.A.; Zhou, X.; Zhu, X.; Dong, Z.; Guo, W. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *Crop J.* **2016**, *4*, 212–219. [[CrossRef](#)]
71. Zhang, J.; Okin, G.S.; Zhou, B. Assimilating optical satellite remote sensing images and field data to predict surface indicators in the Western U.S.: Assessing error in satellite predictions based on large geographical datasets with the use of machine learning. *Remote Sens. Environ.* **2019**, *233*, 111382. [[CrossRef](#)]
72. Sakamoto, T. Incorporating environmental variables into a MODIS-based crop yield estimation method for United States corn and soybeans through the use of a random forest regression algorithm. *ISPRS. J. Photogramm.* **2020**, *160*, 208–228. [[CrossRef](#)]
73. Lobell, D.B.; Asner, G.P. Climate and management contributions to recent trends in U.S. agricultural yields. *Science* **2003**, *299*, 1032. [[CrossRef](#)]
74. Lobell, D.B.; Burke, M.B.; Tebaldi, C.; Mastrandrea, M.D.; Falcon, W.P.; Naylor, R.L. Prioritizing climate change adaptation needs for food security in 2030. *Science* **2008**, *319*, 607–610. [[CrossRef](#)]
75. Tao, F.; Zhang, Z.; Shi, W.; Liu, Y.; Xiao, D.; Zhang, S.; Zhu, Z.; Wang, M.; Liu, F. Single rice growth period was prolonged by cultivars shifts, but yield was damaged by climate change during 1981–2009 in China, and late rice was just opposite. *Glob. Change Biol.* **2013**, *19*, 3200–3209. [[CrossRef](#)]
76. Tao, F.; Zhang, S.; Zhang, Z.; Rotter, R.P. Maize growing duration was prolonged across China in the past three decades under the combined effects of temperature, agronomic management, and cultivar shift. *Glob. Change Biol.* **2014**, *20*, 3686–3699. [[CrossRef](#)]
77. Liu, B.; Asseng, S.; Müller, C.; Ewert, F.; Elliott, J.; David, B.L.; Martre, P.; Ruane, A.C.; Wallach, D.; Jones, J.W.; et al. Similar estimates of temperature impacts on global wheat yield by three independent methods. *Nat. Clim. Chang.* **2016**, *6*, 1130–1136. [[CrossRef](#)]
78. Fletcher, A.L.; Chen, C.; Ota, N.; Lawes, R.A.; Oliver, Y.M. Has historic climate change affected the spatial distribution of water-limited wheat yield across Western Australia? *Clim. Chang.* **2020**, *159*, 347–364. [[CrossRef](#)]
79. Ye, Z.; Qiu, X.; Chen, J.; Cammarano, D.; Ge, Z.; Ruane, A.C.; Liu, L.; Tang, L.; Cao, W.; Liu, B.; et al. Impacts of 1.5 °C and 2.0 °C global warming above pre-industrial on potential winter wheat production of China. *Eur. J. Agron.* **2020**, *120*, 126149. [[CrossRef](#)]
80. Salman, S.A.; Shahid, S.; Sharafati, A.; Salem, G.S.A.; Bakar, A.A.; Farooque, A.A.; Chung, E.-S.; Ahmed, Y.A.; Mikhail, B.; Yaseen, Z.M. Projection of Agricultural Water Stress for Climate Change Scenarios: A Regional Case Study of Iraq. *Agriculture* **2021**, *11*, 1288. [[CrossRef](#)]
81. Mall, R.K.; Lal, M.; Bhatia, V.S.; Rathore, L.S.; Singh, R. Mitigating climate change impact on soybean productivity in India: A simulation study. *Agric. Forest Meteorol.* **2004**, *121*, 113–125. [[CrossRef](#)]
82. Wang, P.; Wu, D.; Yang, J.; Ma, Y.; Feng, R.; Huo, Z. Summer maize growth under different precipitation years in the Huang-Huai-Hai Plain of China. *Agric. Forest Meteorol.* **2020**, *285*, 107927. [[CrossRef](#)]
83. Li, J.; Lei, H. Impacts of climate change on winter wheat and summer maize dual-cropping system in the North China Plain. *Environ. Res. Commun.* **2022**, *4*, 075014. [[CrossRef](#)]
84. Cao, Y.; Qi, J.; Wang, F.; Li, L.; Lu, J. Analysis of climate suitability of spring maize in Liaoning Province based on modulus and mathematics. *Sci. Geogr. Sin.* **2020**, *40*, 1210–1220.
85. Liu, B.; Liu, L.; Asseng, S.; Zhang, D.; Ma, W.; Tang, L.; Cao, W.; Zhu, Y. Modelling the effects of post-heading heat stress on biomass partitioning, and grain number and weight of wheat. *J. Exp. Bot.* **2020**, *71*, 6015–6031. [[CrossRef](#)]
86. Song, X.; Zhang, Z.; Chen, Y.; Wang, P.; Xiang, M.; Shi, P.; Tao, F. Spatiotemporal changes of global extreme temperature events (ETEs) since 1981 and the meteorological causes. *Nat. Hazards* **2013**, *70*, 975–994. [[CrossRef](#)]
87. Zhu, X.; Liu, T.; Xu, K.; Chen, C. The impact of high temperature and drought stress on the yield of major staple crops in northern China. *J. Environ. Manag.* **2022**, *314*, 115092. [[CrossRef](#)] [[PubMed](#)]
88. Chen, R.; Wang, J.; Li, Y.; Song, Y.; Huang, M.; Feng, P.; Qu, Z.; Liu, L. Quantifying the impact of frost damage during flowering on apple yield in Shaanxi province, China. *Eur. J. Agron.* **2023**, *142*, 126642. [[CrossRef](#)]

89. Liu, B.; Liu, L.; Tian, L.; Cao, W.; Zhu, Y.; Asseng, S. Post-heading heat stress and yield impact in winter wheat of China. *Glob. Chang. Biol.* **2014**, *20*, 372–381. [[CrossRef](#)] [[PubMed](#)]
90. Xiao, L.; Liu, B.; Zhang, H.; Gu, J.; Fu, T.; Asseng, S.; Liu, L.; Tang, L.; Cao, W.; Zhu, Y. Modeling the response of winter wheat phenology to low temperature stress at elongation and booting stages. *Agric. Forest Meteorol.* **2021**, *303*, 108376. [[CrossRef](#)]
91. Xiao, L.; Liu, L.; Asseng, S.; Xia, Y.; Tang, L.; Liu, B.; Cao, W.; Zhu, Y. Estimating spring frost and its impact on yield across winter wheat in China. *Agric. Forest Meteorol.* **2018**, *260*, 154–164. [[CrossRef](#)]
92. Lobell, D.B.; Roberts, M.J.; Schlenker, W.; Braun, N.; Little, B.B.; Rejesus, R.M.; Hammer, G.L. Greater sensitivity to drought accompanies maize yield increase in the U.S. Midwest. *Science* **2014**, *344*, 516–519. [[CrossRef](#)]
93. Zhao, C.; Liu, B.; Piao, S.; Wang, X.; Lobell, D.B.; Huang, Y.; Huang, M.; Yao, Y.; Bassu, S.; Ciais, P.; et al. Temperature increase reduces global yields of major crops in four independent estimates. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 9326–9331. [[CrossRef](#)]
94. Bitu, C.E.; Gerats, T. Plant tolerance to high temperature in a changing environment: Scientific fundamentals and production of heat stress-tolerant crops. *Front. Plant Sci.* **2013**, *4*, 273. [[CrossRef](#)]
95. Wang, N.; Wang, E.; Wang, J.; Zhang, J.; Zheng, B.; Huang, Y.; Tan, M. Modelling maize phenology, biomass growth and yield under contrasting temperature conditions. *Agric. Forest Meteorol.* **2018**, *250*, 319–329. [[CrossRef](#)]
96. Siddik, M.A.; Zhang, J.; Chen, J.; Qian, H.; Jiang, Y.; Raheem, A.K.; Deng, A.; Song, Z.; Zheng, C.; Zhang, W. Responses of indica rice yield and quality to extreme high and low temperatures during the reproductive period. *Eur. J. Agron.* **2019**, *106*, 30–38. [[CrossRef](#)]
97. You, L.; Rosegrant, M.W.; Wood, S.; Sun, D. Impact of growing season temperature on wheat productivity in China. *Agric. Forest Meteorol.* **2009**, *149*, 1009–1014. [[CrossRef](#)]
98. Bai, H.; Xiao, D.; Wang, B.; Liu, L.; Tang, J. Simulation of Wheat Response to Future Climate Change Based on Coupled Model Inter-Comparison Project Phase 6 Multi-Model Ensemble Projections in the North China Plain. *Front. Plant Sci.* **2022**, *13*, 829580. [[CrossRef](#)]
99. Liu, B.; Asseng, S.; Wang, A.; Wang, S.; Tang, L.; Cao, W.; Zhu, Y.; Liu, L. Modelling the effects of post-heading heat stress on biomass growth of winter wheat. *Agric. Forest Meteorol.* **2017**, *247*, 476–490. [[CrossRef](#)]
100. Glotter, M.; Elliott, J. Simulating US agriculture in a modern Dust Bowl drought. *Nat. Plants* **2016**, *3*, 16193. [[CrossRef](#)]
101. Xu, C.; McDowell, N.G.; Fisher, R.A.; Wei, L.; Sevanto, S.; Christoffersen, B.O.; Weng, E.; Middleton, R.S. Increasing impacts of extreme droughts on vegetation productivity under climate change. *Nat. Clim. Chang.* **2019**, *9*, 948–953. [[CrossRef](#)]
102. Chiang, F.; Mazdiyasn, O.; AghaKouchak, A. Evidence of anthropogenic impacts on global drought frequency, duration, and intensity. *Nat. Commun.* **2021**, *12*, 2754. [[CrossRef](#)]
103. Kreibich, H.; Van Loon, A.F.; Schroter, K.; Ward, P.J.; Mazzoleni, M.; Sairam, N.; Abeshu, G.W.; Agafonova, S.; AghaKouchak, A.; Aksoy, H.; et al. The challenge of unprecedented floods and droughts in risk management. *Nature* **2022**, *608*, 80–86. [[CrossRef](#)]
104. Ma, Y.; Zhang, Z.; Kang, Y.; Özdoğan, M. Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. *Remote Sens. Environ.* **2021**, *259*, 112408. [[CrossRef](#)]
105. Lobell, D.B.; Thau, D.; Seifert, C.; Engle, E.; Little, B. A scalable satellite-based crop yield mapper. *Remote Sens. Environ.* **2015**, *164*, 324–333. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.