Article

Optimized Random Forest Framework for Integrating Cultivar, Environmental, and Phenological Interactions in Crop Yield Prediction

Jiaojiao Tan [1,2,3], Lu Jiang [4], Yingnan Wei [4], Ning Yao [5,*], Gang Zhao [3] and Qiang Yu [3]

1   The Research Center of Soil and Water Conservation and Ecological Environment, Chinese Academy of Sciences and Ministry of Education, Yangling 712100, China; tanjiaojiao19@mails.ucas.ac.cn
2   Institute of Soil and Water Conservation, Chinese Academy of Sciences and Ministry of Water Resources, Yangling 712100, China
3   State Key Laboratory of Soil and Water Conservation and Desertification Control, Northwest A&F University, Yangling 712100, China
4   College of Soil and Water Conservation Science and Engineering, Northwest A&F University, Yangling 712100, China
5   College of Water Resources and Architectural Engineering/Key Lab of Agricultural Soil and Water Engineering in Arid and Semiarid Areas, Ministry of Education, Northwest A&F University, Yangling 712100, China
*   Correspondence: yaoning@nwafu.edu.cn

Abstract

Accurate rice yield prediction remains a major challenge due to the complex and nonlinear interactions among cultivar, environment, and phenology. Existing approaches often focus on analyzing individual components while ignoring their interdependencies, which results in limited predictive accuracy and generalizability. To overcome these problems, this study proposes a novel interpretable random forest model that integrates cultivar, environmental, and phenological dimensions. Different from conventional approaches, the proposed method incorporates a factor-combination optimization strategy to identify the most effective information for yield estimation. For analysis, 24 key determinants were screened, including the geographical location, meteorological conditions, phenological events, and cultivar traits. The RF models were also evaluated when built with seven factor combinations. The results reveal the following: (1) Meteorological conditions play a dominant role during the vegetative growth period, including net solar radiation ($r = 0.42$), daylength ($r = 0.38$), and thermal summation ($r = 0.29$). On the other hand, thermal summation ($r = 0.28$), mean minimum temperature ($r = -0.23$), and mean temperature ($r = -0.20$) are most relevant during the reproductive growth period. (2) The full-factor model achieves optimal performance (RMSE = 601.45 kg/ha and MAE = 454.98 kg/ha, $R^2 = 0.77$). (3) Importance analysis reveals that meteorological factors provide the greatest contribution (53.59%), followed by phenological factors (20.39%), geographical factors (17.20%), and cultivar (8.82%), respectively. The results also reveal that threshold effects of key determinants on yield, and identify mid-April to early May as the optimal sowing window. These findings demonstrate that integrating cultivar, environment, and phenology factors creates a powerful predictive model for rice yields.

Keywords: yield prediction; random forest; rice cultivar; environment; phenology

1. Introduction

Rice is one of the world's most vital staple crops, playing a central role in ensuring global food security [1]. It provides a primary source of calories and nutrition for billions of people worldwide. However, the stability and improvement in rice yields are increasingly threatened by accelerating climate change, more frequent extreme weather events, and the continual replacement of rice cultivars [2–6]. These challenges create complex risks for production systems, necessitating more precise and adaptive management strategies to sustain yield stability amid changing environmental conditions.

Among the many factors influencing rice yield formation, the complex and nonlinear interactions between cultivar genetics, environmental conditions, and phenological development are especially important. Together, these factors govern the rate and pattern of rice growth, fundamentally determining both rice's potential yield and actual outcomes in the field. Key environmental variables, such as temperature patterns, solar radiation intensity and duration, and water availability—as well as their dynamic temporal distributions—strongly affect critical developmental stages, including tillering, heading, flowering, and grain filling [7–12]. For instance, photoperiod length and accumulated thermal time during crucial growth periods regulate the timing of heading and grain filling, which are essential to both yield quantity and quality [13–15]. Extreme temperatures, whether excessively high or low, can severely disrupt panicle formation, impair pollination, and reduce grain weight and quality, thereby significantly lowering final yield [16–18].

In parallel, genetic diversity among rice cultivars results in considerable variation in their sensitivity and adaptability to environmental stresses [19]. Early- and late-maturing varieties often differ markedly in photoperiod sensitivity, optimal temperature ranges, and developmental duration [20–22]. Early-maturing cultivars tend to have weaker photoperiod sensitivity and are less affected by daylength variations, whereas late-maturing cultivars generally exhibit stronger photoperiodic responses—with short days encouraging heading, and long days delaying it. Achieving optimal alignment between genotype and environment is therefore critical; the precise control of phenological development enables sensitive growth stages to coincide with favorable climatic windows, maximizing yield and stability.

Despite decades of research on rice yield prediction, current methodologies continue to face significant limitations. Traditional statistical models, often relying on historical datasets and employing univariate or simple multivariate regression techniques, struggle to capture the complex, nonlinear interactions inherent in crop growth systems [23–25]. Furthermore, these models lack robustness when applied to new or changing environmental conditions. Process-based mechanistic crop models offer greater biological interpretability by simulating key physiological processes, such as photosynthesis, respiration, and assimilate partitioning [26–30]. However, these models encounter challenges, including difficulties in accurately calibrating parameters across diverse regions, high sensitivity to uncertain inputs, limited ability to integrate heterogeneous data sources, and constraints in adapting to new cultivars or evolving management practices [31–34]. Importantly, many existing approaches tend to overlook the intricate interactions among cultivar genetics, environment, and crop management that are critical for comprehensive yield prediction. In particular, the dynamic coupling between cultivar-dependent management practices and their yield-regulating mechanisms has received little attention.

To address these challenges, this study proposes an interpretable, data-driven modeling framework for rice yield prediction that systematically integrates multidimensional factors. Leveraging a rich observational dataset covering 2175 rice cultivars tested across 327 field trial sites in China's major rice-producing regions from 2007 to 2018, this study incorporates 47 variables spanning meteorological, geographical, phenological, and cultivar-related

categories. The specific objectives are to (1) identify the key determinants of rice yield formation using multiple factors and thoroughly investigating the differentiated response mechanisms of rice yield to various factors; (2) develop seven random forest regression models with different feature combinations to evaluate predictive performance and determine the optimal model; and (3) quantitatively assess the contribution of each input variable to predictive accuracy and elucidate the marginal effects of these variables. These findings provide a scientifically grounded foundation for precision agriculture, resource optimization, and the development of climate-resilient cultivation practices. Ultimately, this integrative approach advances the mechanistic understanding of rice yield formation and supports efforts to enhance yield stability and food security in the face of increasing climate variability and environmental challenges.

2. Materials and Methods

2.1. Study Area and Data Sources

The study area encompasses China's rice cultivation regions, with trial sites distributed as shown in Figure 1. These sites span six major rice-producing zones across China, characterized by pronounced heterogeneity in edaphoclimatic conditions, socioeconomic conditions, and rice cultivation systems. Those zones are the South China double-season rice region, Central China mixed single-season and double-season rice region, Southwest Plateau mixed single-season and double-season rice region, North China single-season rice region, Northeast single-season rice region, and Northwest arid single-season rice region. The study area domain extends latitudinally from 19°09′ N to 46°40′ N and longitudinally from 80°07′ E to 130°30′ E, with elevational gradients ranging from 1 to 1318 m. This extensive geographical expanse encompasses diverse climatic conditions spanning tropical, subtropical, and temperate bioclimatic zones. During the main rice growing season (March–November), a significant spatiotemporal fluctuation in meteorological conditions is observed. The climate conditions across the study area are highly heterogeneous, with multi-year daily mean temperature, daily maximum temperature, and daily minimum temperature during the growing season at all trial sites varying from 9.0 °C to 27.1 °C, 13.8 °C to 29.7 °C, and 4.6 °C to 25.3 °C, respectively; accumulated rainfall varying from 25 mm to 2812 mm; average daily relative humidity varying from 26.82% to 84.15%; and accumulated net solar radiation varying from 2482 MJ·m^{-2}·d^{-1} to 4622 MJ·m^{-2}·d^{-1}.
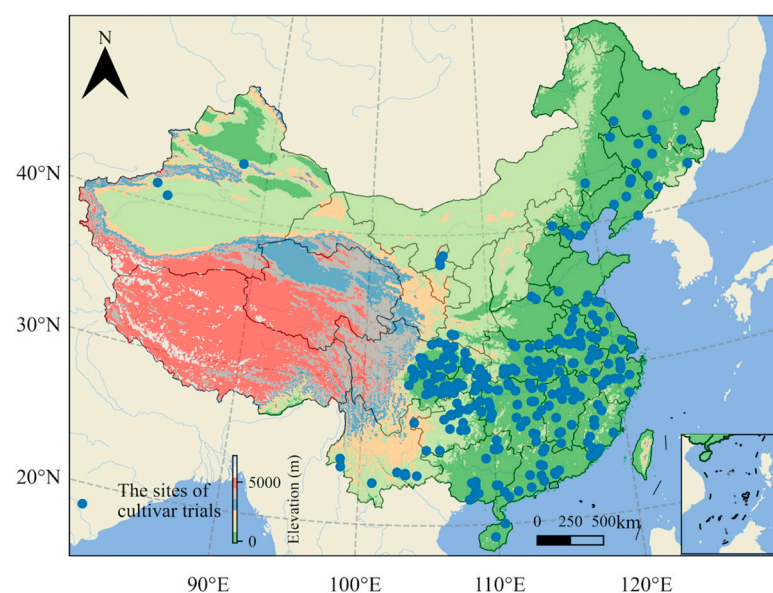

Figure 1. The study area and the distribution of sites for rice cultivar trials.

2.1.1. Rice Cultivar Trials Data

(1)　Data source

The rice cultivar trials data in this study are obtained from the "China Rice New Cultivar Trials" series published by China Agricultural Science and Technology Press [35]. The series of authoritative publications comprehensively document regional production and field plot trials of rice cultivars across China from 2007 to 2018. Field plot trials are conducted at representative sites within specific ecological zones that reflect local edaphic conditions, climatic patterns, agronomic practices, and productivity levels to evaluate varietal adaptability. These plot trials are conducted in a completely randomized block design with triplicate repetitions for each cultivar on 13.33 m^2. Regional production trials, conducted as a secondary evaluation phase, verify the yield performance of field plot trials. These production trials utilize randomized field arrangements without repetition for each hybrid cultivar on approximately 333.34 m^2. Each volume in these books provides detailed documentation, including geographical location, participating germplasm (cultivar, male and female parents), phenological date, phenotypic traits, disease resistance, yield, and cultivar traits for each trial.

This study collected data on 2175 rice cultivars with a total of 46,293 observations at 327 locations from 2007 to 2018. The data include the geographic location of trial sites, such as longitude, latitude, and elevation; yield and yield components, such as the number of effective panicles per unit area (EPPA), total number of grains per panicle (TGPP), filled grains per panicle (FGPP), seed-setting rate (SSR), and thousand-grain weight (TGW); phenology, such as sowing date, transplanting date, heading date, and maturation date; and cultivar traits, such as plant height (PH) and panicle length (PL). The results with large errors and affected by natural disasters, pests, and diseases are not included in the statistical summary to ensure data accuracy and reliability. All variable abbreviations and descriptions can be found in Table A1.

(2)　The data distribution of rice cultivar trials

According to observation records, the dates of the occurrence of the same phenological event at different trial sites vary significantly. Specifically, sowing occurs between February and July, heading occurs between April and October, and maturation occurs between May and November. The day of year for sowing date (DOY_Sow) ranges from day 33 to 207 (mean = 119.6 ± 34.7 days), the day of year for heading date (DOY_Hea) ranges from day 114 to 293 (mean = 220.5 ± 28.3 days), and the day of year for maturation date (DOY_Mat) ranges from day 142 to 331 (mean = 256.2 ± 31.9 days). The entire development period (GP) from sowing to maturation averages 137.6 ± 17.3 days (range: 89–194 days), with the vegetative growth period (VGP) from sowing to heading averaging 100.9 ± 16.0 days, and the reproductive growth period (RGP) from heading to maturation averaging 36.7 ± 6.8 days.

Statistics show that among 46,293 samples, the yield data are complete, but there are some missing data on seven cultivar-related traits. Missing data rate ranges from 11.48% to 13.53% across seven traits, with 88.19% of samples having complete data for all traits. Figure 2 provides the violin plots to show a visual representation of the data distribution of rice cultivar trials (Table A2 provides the statistical distribution for all variables incorporated in the model development). Rice yield averages 8664.00 ± 1290.00 kg/ha (range: 5162.55–1.22 × 10^4 kg/ha), demonstrating significant variability. Among yield components, EPPA averages 258.30 ± 55.2 × 10^4 per ha (range: 118.50–838.50 × 10^4 per ha), TGPP averages 170.8 ± 41.1 (range: 51.6–518.8 gains per panicle), and FGPP averages 138.5 ± 33.0 (range: 20.8–369.8 gains per panicle). The SSR averages 81.26 ± 8.24% (range: 19.00–100%), while TGW averages 27.03 ± 2.94 g (range: 16.20–38.70 g). For other traits, PH averages 112.56 ± 14.02 cm (range: 18.00–202.00 cm), and PL averages 23.70 ± 2.93 cm

(range: 11.20–38.00 cm). This dataset provides a comprehensive foundation for exploring the relationships between phenological development, cultivar traits, and yield formation across diverse cultivars and growing conditions.
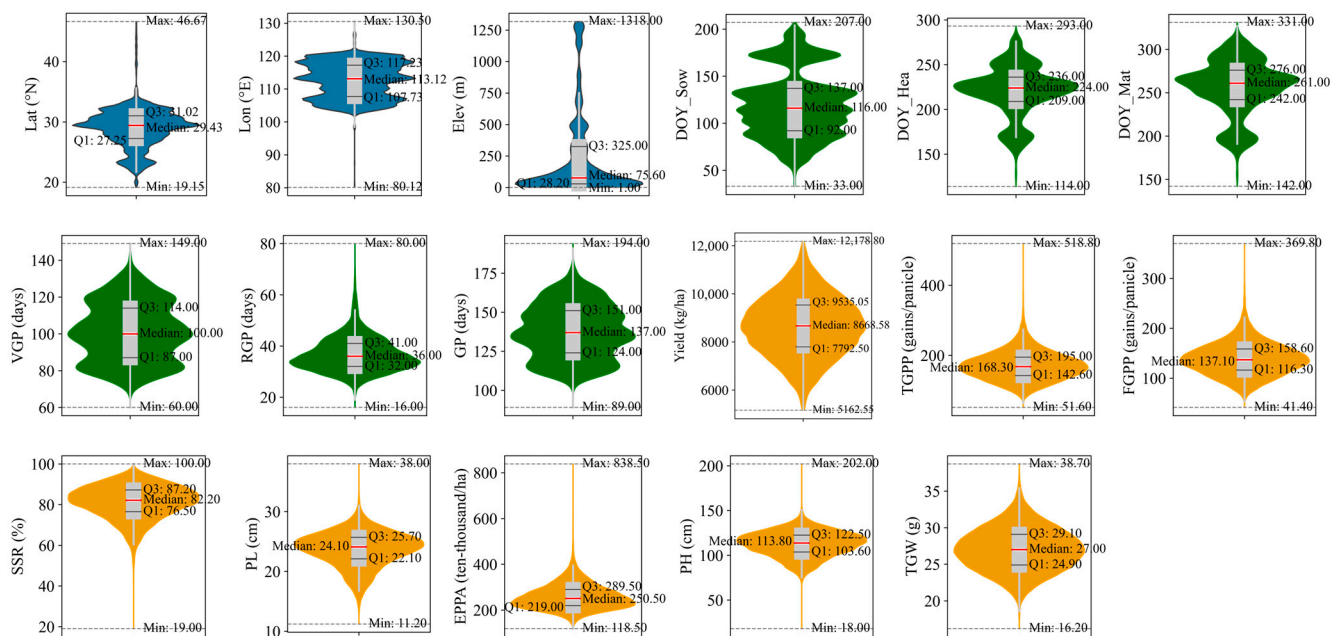

Figure 2. Visualization of the variable distribution in rice cultivar trials dataset. Variables are categorized and indicated by different colors—blue (geographical factors), green (phenological factors), and orange (cultivar traits).

2.1.2. Meteorological Data

The meteorological data matching the new Chinese rice cultivar trials is collected from the ERA5-Land (https://cds.climate.copernicus.eu/datasets (accessed on 19 September 2025)) dataset, which is a high-resolution global atmospheric reanalysis dataset developed and maintained by the European Centre for Medium-Range Weather Forecasts (ECMWF), providing continuous, consistent, and high-quality data for various meteorological variables. This study extracts daily meteorological data from 2007 to 2018 according to the latitude and longitude of trial sites. Then, the units of each meteorological variable are converted to determine the daily minimum temperature (TMin, °C), daily maximum temperature (TMax, °C), daily mean temperature (TMean, °C), daily precipitation (PRE, mm), daily relative humidity (RH, %), and daily net solar radiation (Rns, MJ·m^{-2}·d^{-1}). The daylength (DL, h) for a specific location and date was calculated based on longitude, latitude, and the day of year [36]. The thermal summation during the developmental period is also calculated, such as ≥ 8 °C thermal summation (TS, °C d). For the entire development period, a series of meteorological stress indicators are calculated, including the frequency of high-temperature events (3 consecutive days with daily average temperature ≥ 30 °C, HN), total days with high-temperature events (HD), accumulated heat of HD (HDD), the frequency of cold damage events (daily mean temperature ≤ 17 °C above 36° N or below 20 °C below 36° N for consecutive 3 days, CN), total days with cold-damage events (CD), accumulated cold of CD (CDD, °C d), and total days with high-heat and high-humidity (daily average temperature ≥ 25.0 °C and relative humidity $\geq 90.0\%$, HHD). All variable abbreviations and descriptions can be found in Table A1.

2.2. Research Methods

2.2.1. Descriptive Statistical Analysis

Descriptive statistical analysis is used to summarize the essential features of a dataset, capturing both its center and its spread. The mean pinpoints central tendency, while the standard deviation (Std) quantifies dispersion. The mean is the average of all samples in a dataset, calculated as

$$Mean = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

where x_i is the *i*-th value of the variable in the dataset, and *n* is the sample size of the variable in the dataset.

Standard deviation (S_{td}) measures the dispersion of variable values in a dataset, representing the average deviation of each variable value from its mean, calculated as

$$Std = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - Mean)^2} \tag{2}$$

2.2.2. Correlation Analysis

Spearman's rank correlation coefficient is used to measure the monotonic relationship between two variables, which is a nonparametric statistical method. It calculates correlation by converting original data to ranks (positions after sorting), does not depend on the distribution form of the data, and can effectively handle nonlinear relationships or non-normally distributed data. Since not all variables follow normal distributions, Spearman's rank correlation coefficient is leveraged to evaluate the correlation between yield and each influencing variable and generate correlation matrix heatmaps and data distribution plots. The Spearman correlation coefficient is calculated as

$$r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)} \tag{3}$$

where d_i is the rank difference of each pair of statistical values, and *n* is the sample size of the statistical variables.

2.2.3. Model Evaluation Methods

Three statistical indicators are used to evaluate model accuracy, including mean absolute error (MAE, Equation (4)), root mean square error (RMSE, Equation (5)), and coefficient of determination (R^2, Equation (6)). The closer MAE and RMSE are to 0, the smaller the error; the closer R^2 is to 1, the better the model fit.

The calculation methods for MAE, RMSE, and R^2 are as follows:

$$MAE = \sum_{i=1}^{n} |S_i - o_i| / n \tag{4}$$

$$RMSE = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n} (S_i - O_i)\right)} \tag{5}$$

$$R^2 = \frac{\left(\sum_{i=1}^{n} (S_i - \overline{S})(O_i - \overline{O})\right)^2}{\sum_{i-1}^{n} (S_i - \overline{S})^2 \sum_{i=1}^{n} (O_i - \overline{O})^2} \tag{6}$$

where O_i is the *i*-th observed value, S_i is the *i*-th predicted value, *n* is the sample size of the observed variables, $\overline{O}$ is the mean of observed values, and $\overline{S}$ is the mean of predicted values.

2.2.4. Development of a Random Forest-Based Yield Prediction Model

The random forest regression algorithm, as a powerful nonparametric statistical method, is used to handle complex nonlinear relationships of multidimensional data to predict crop yields. In the data preprocessing phase, the interquartile range (IQR) method is utilized to identify and remove yield outliers, specifically eliminating values below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$ to ensure dataset integrity and reliability. Additionally, the 2175 rice cultivars are categorical variables (fields) rather than quantitative variables (numerical values). To enable the model to properly handle this categorical feature within the analytical framework, each cultivar is encoded with a unique identifier. The approach allows the model to capture cultivar-specific effects without imposing any numerical ordering or relationship between different cultivars.

When selecting feature variables for random forest input, this study covers geographic location information, phenological factors, meteorological factors, and cultivar traits. Then, the min–max normalization method is used to standardize feature variables, making the data distribution of each feature on the same scale. This processing not only helps improve model convergence speed but also avoids adverse effects on model weight allocation caused by large differences in different feature scales. To evaluate the impact of different feature combinations on crop yield prediction accuracy, the proposed method designs a series of experiments, gradually adding variables. Crop yield consistently serves as the target variable throughout all experiments. The specific steps are as follows:

First, the method conducts single-dimension feature experiments: the first experiment uses only geographical location (Loc) as the input feature, the second employs only a phenological factor (Phen), and the third isolates the meteorological factor (Meteo). Next, two-dimensional feature combinations are progressed: the fourth integrates Phen with Meteo, while the fifth combines Loc with Meteo. Then, the paper advances to three-dimensional feature integration in the sixth experiment, incorporating Loc, Phen, and Meteo simultaneously. Finally, a comprehensive multi-dimensional analysis in the seventh experiment is implemented by introducing cultivar traits, using the full set of features—Loc, Phen, Meteo, and cultivar traits (please refer to Table A1 for the specific variables of the input models).

For each model, the method optimizes random forest regressor hyperparameters using grid search with five-fold cross-validation. The hyperparameters include tree count, maximum depth, minimum samples for node splitting, minimum samples in leaf nodes, and maximum feature count. The coefficient of determination (R^2) is used to determine the optimal parameter combination, ensuring optimal model performance.

In both model training and evaluation phases, this study employs a training–testing split method with 100 iterations and implements parallel computing to improve efficiency. In each iteration, the dataset is stratified by geographical coordinates (Lon and Lat) and randomly divided into an 80% training set and a 20% testing set. This stratification ensures balanced geographical distribution between sets, reducing risks of overfitting or underfitting due to geographical disparities. During each training–testing process, RMSE, MAE, and R^2 are recorded to assess model accuracy and generalization capability.

The result is ultimately presented as the mean performance metrics across all 100 iterations for both training and testing sets. These statistics can demonstrate the random forest regression model's stability and reliability in crop yield prediction, providing a foundation for agricultural decision making and yield prediction research.

The workflow of the proposed methods and the data and models used in this study are shown in Figure 3.
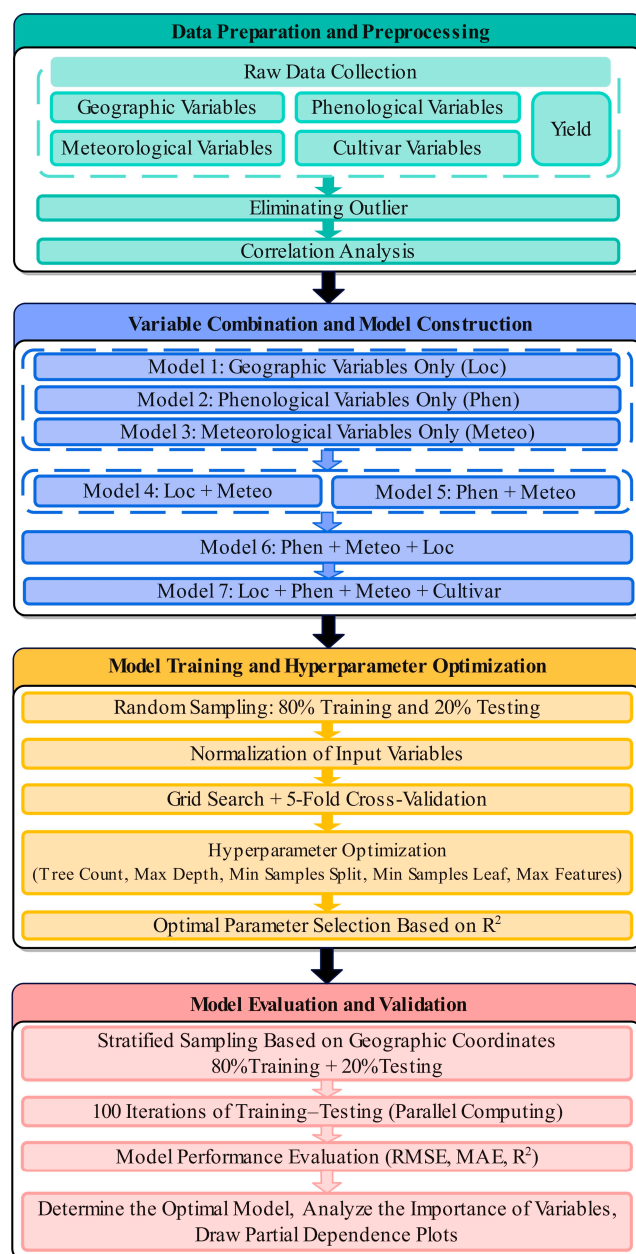
Figure 3. The workflow of the proposed method. Refer to Table A1 for the specific variables of the input models.

3. Results

3.1. Correlation Analysis Between Influencing Factors and Rice Yield

3.1.1. The Correlation Between Geography, Phenology, and Rice Yield

The influence of trial geographic location, key phenological dates, and the durations of different development periods on rice yield is comprehensively considered. Geographic location information includes longitude (Lon), latitude (Lat), and elevation (Elev); phenological factors include DOY_Sow, DOY_Hea, VGP, RGP, and GP. The relationships between these variables and yield are shown in Figure 4a.
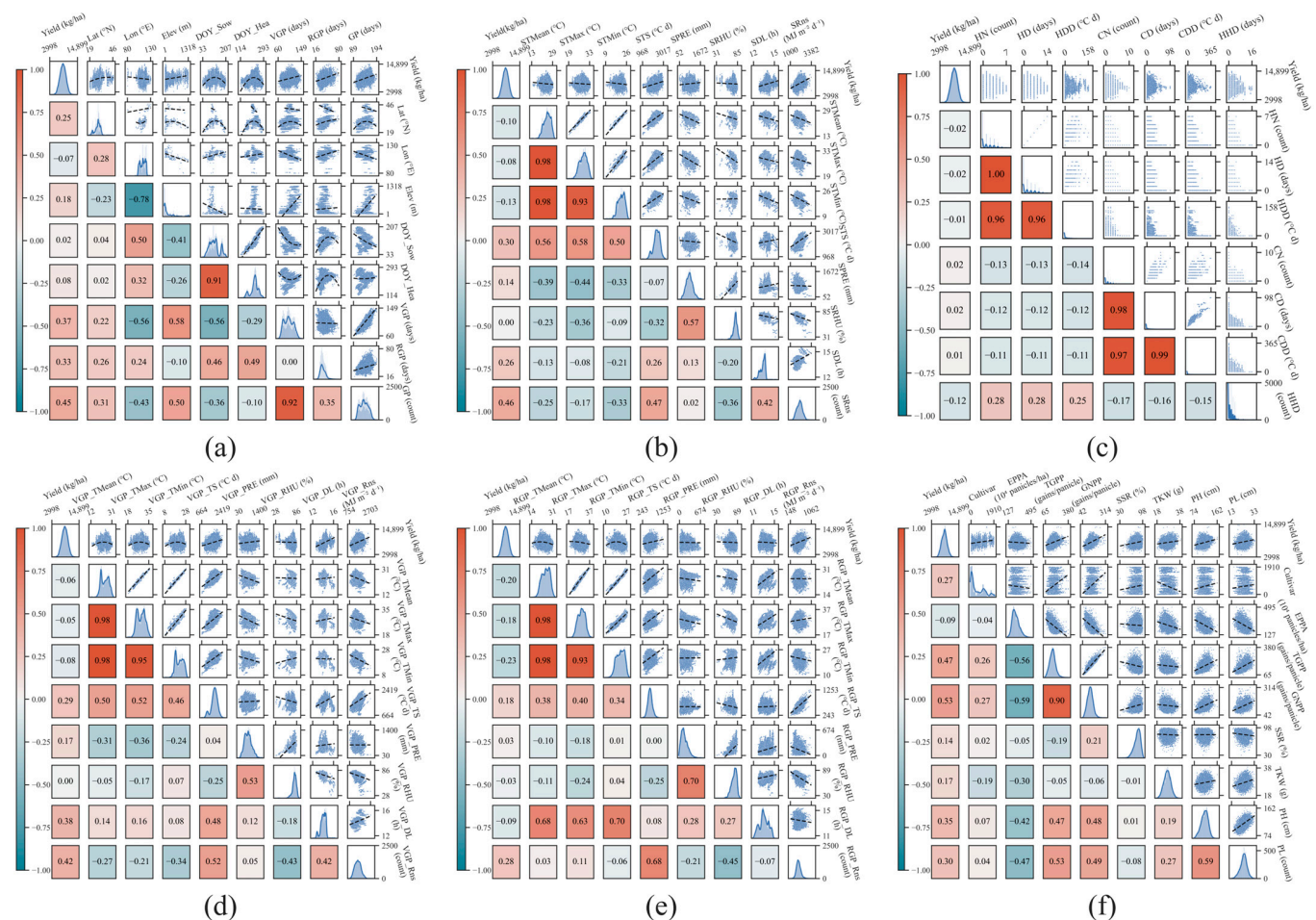
Figure 4. Correlation analysis between influencing factors and rice yield. Panels show correlations of (**a**) geographical location, phenology, and yield; (**b**) meteorological factors during the whole growth period (GP); (**c**) meteorological stress indicators during GP; (**d**) meteorological factors during the vegetative growth phase (VGP); (**e**) meteorological factors during the reproductive growth phase (RGP); and (**f**) cultivar traits. Diagonal plots present histograms and kernel density estimates. Variables include location (Lon, Lat, and Elev), phenology (DOY_Sow, DOY_Hea, VGP, RGP, and GP), meteorological factors (TMin, TMax, TMean, PRE, RHU, Rns, DL, and TS), stress indicators (HN, HD, HDD, CN, CD, CDD, and HHD), and cultivar traits (EPPA, TGPP, FGPP, SSR, TGW, PH, and PL). The dashed lines represent the fitting lines of the data points. All variable abbreviations and descriptions can be found in Table A1.

Lat and Elev have significant linear correlations with rice yield, with correlation coefficients of 0.25 and 0.18, respectively. The correlation between yield and Lon is weaker, at -0.07. VGP, RGP, and GP are significantly positively correlated with yield, with correlation coefficients of 0.37, 0.33, and 0.45, respectively. The linear correlations between yield and DOY_Sow and DOY_Hea are both weak, with correlation coefficients of 0.02 and 0.08, respectively. But according to the regression plots in the upper triangle, DOY_Sow and DOY_Hea have significant nonlinear relationships with yield.

The correlation analysis between independent variables indicates that there is an interaction effect among the variables. Lon and Elev are significantly negatively correlated ($r = -0.78$, $p < 0.01$). Lat has significant positive correlations with the durations of development periods, with correlation coefficients with VGP, RGP, and GP of 0.22, 0.26, and 0.31, respectively. DOY_Sow and DOY_Hea are significantly influenced by Elev and Lon, with significant negative correlations with Elev (DOY_Sow, $r = -0.41$; DOY_Hea, $r = -0.26$; $p < 0.01$) and significant positive correlations with Lon (DOY_Sow, $r = 0.50$; DOY_Hea,

r = 0.32; *p* < 0.01). The durations of development periods are significantly influenced by the geographic location of trial sites, with VGP, RGP, and GP lengthening with increasing Lat and Elev; they are also influenced by DOY_Sow and DOY_Hea. VGP and GP tend to shorten with delayed sowing date and heading date, and RGP tends to lengthen with delayed sowing date and heading date.

3.1.2. The Correlation Between Rice Yield and Meteorology During Different Development Periods

The impact of varying meteorological conditions on rice development processes and ultimate yield shows significant variability. Therefore, this work separately analyzes the relationships between meteorological factors and yield during different development periods, as shown in Figure 4b–e.

Throughout the development period, rice yield shows strong positive correlations with key meteorological factors: net solar radiation (SRns, r = 0.46), ≥8 °C thermal summation (STS, r = 0.30), daylength (SDL, r = 0.26), and accumulated precipitation (SPRE, r = 0.14).

In contrast, temperature-related variables—mean minimum temperature (STMin), mean temperature (STMean), and mean maximum temperature (STMax)—exert weaker effects, with absolute correlations staying below 0.13. These temperature variables are negatively correlated with moisture indicators (SPRE and SRHU), with coefficients ranging from −0.23 to −0.44; SRHU and SRns also display a moderate negative relationship (r = −0.36). These patterns reveal complex interactions among meteorological factors that jointly shape the growing environment while exerting both independent and combined influences on yield. Furthermore, high-temperature heat damage (HN, HD), low-temperature cold damage (CN, CD), and hot-humid conditions (HHD) occur only rarely across the entire period (Figure 4c). Excluding HHD, HN, HD, CN, and CD are characterized by limited data, scattered distributions, and weak correlations with yield (| r | ≤ 0.02).

During VGP, rice yield correlates most strongly with net solar radiation (VGP_Rns), daylength (VGP_DL), ≥8 °C thermal summation (VGP_TS), and accumulated precipitation (VGP_PRE), with correlation coefficients of 0.42, 0.38, 0.29, and 0.17, respectively. Temperature variables (VGP_TMean, VGP_TMax, and VGP_TMin) exhibit weak negative correlations with yield (−0.08 ≤ r < 0). In contrast, during RGP, yield is influenced primarily by net solar radiation (RGP_Rns, r = 0.28), mean minimum temperature (RGP_TMin, r = −0.23), mean temperature (RGP_TMean, r = −0.20), mean maximum temperature (RGP_TMax, r = −0.18), and ≥8 °C thermal summation (RGP_TS, r = 0.18). Accumulated precipitation (RGP_PRE) and mean relative humidity (RGP_RHU) show comparatively weaker associations with final yield.

Taken together, no single meteorological variable fully explains yield variation. The interactions among multiple environmental factors must be considered. During VGP, VGP_Rns, VGP_TS, VGP_DL, and VGP_PRE exert the strongest effects; during RGP, yield is more related to RGP_Rns, RGP_TMin, and RGP_TMean. Across the entire development period, HHD also negatively affects yield. Consequently, the developmental period is partitioned into two distinct phases: VGP and RGP. All meteorological variables except VGP_RHU are adopted as model inputs, and among the stress indicators, only HHD is selected.

3.1.3. The Correlation Between Cultivar Traits and Rice Yield

This study examines the relationships between rice yield and cultivar traits, as shown in Figure 4f. FGPP and TGPP exhibit strong positive correlation with yield (FGPP, r = 0.51; TGPP, r = 0.45) and are themselves highly intercorrelated (r = 0.90). PH and PL also correlate positively with yield (r = 0.37 and r = 0.27, respectively). And the cultivar itself displays a positive association with yield (r = 0.27). In contrast, TGW, SSR, and EPPA exhibit weaker

correlations with yield (r = 0.16, r = 0.14, and r = −0.08, respectively). Importantly, yield components are subject to mutual constraints: TGW is negatively correlated with EPPA (r = −0.31), and TGPP is negatively correlated with SSR (r = −0.21). Further analysis reveals that cultivar is positively correlated with both TGPP and FGPP (r > 0.2), underscoring the role of genetic factors in panicle development. TGPP and FGPP also correlate strongly with PH and PL (r ≈ 0.5), indicating that taller plants with longer panicles tend to produce more grains—traits governed by cultivar genetics. Conversely, EPPA is negatively correlated with TGPP and FGPP (r = −0.57 and r = −0.59, respectively), suggesting that excessive tillering can reduce grains per panicle and potentially limit economic yield.

Rice phenotypic traits are strongly influenced by meteorological conditions (Table 1). Temperatures (VGP_TMean, VGP_TMax, and VGP_TMin) during the vegetative growth period can promote EPPA and PH (0.14 ≤ r ≤ 0.17), but reduce TGW (−0.26 ≤ r ≤ −0.24). Precipitation in the vegetative growth period decreases EPPA (r = −0.24), but increases TGPP, FGPP, and TGW (0.13 ≤ r ≤ 0.19). TS during vegetative and reproductive growth periods positively affects TGPP, FGPP, PH, and PL (0.17 ≤ r ≤ 0.47). DL during the vegetative growth period positively affects TGPP, FGPP, PH, and PL (0.26 ≤ r ≤ 0.50). Rns during vegetative and reproductive growth periods also shows significant positive effects on most cultivar traits (0.04 ≤ r ≤ 0.35), except that Rns during the vegetative growth period has negative effects on EPPA (r = −0.26) and SSR (−0.04).

Table 1. The correlation between cultivar traits and meteorological factors.

	EPPA	TGPP	FGPP	SSR	TGW	PH	PL
VGP_TMean	0.16	0.02	0.00	−0.07	−0.25	0.17	0.01
VGP_TMax	0.17	0.01	−0.02	−0.08	−0.24	0.16	0.01
VGP_TMin	0.14	0.03	0.00	−0.08	−0.26	0.14	0.00
VGP_TS	−0.21	0.36	0.31	−0.11	0.07	0.47	0.33
VGP_PRE	−0.24	0.18	0.19	0.01	0.13	0.11	0.14
VGP_RHU	−0.07	0.05	0.07	0.04	−0.09	−0.03	−0.02
VGP_DL	−0.18	0.30	0.34	0.08	0.06	0.50	0.26
VGP_Rns	−0.26	0.31	0.28	−0.04	0.32	0.35	0.34
RGP_TMean	-	-	−0.02	0.12	−0.06	0.08	0.07
RGP_TMax	-	-	−0.01	0.12	−0.06	0.09	0.08
RGP_TMin	-	-	−0.04	0.10	−0.08	0.06	0.05
RGP_TS	-	-	0.32	0.11	−0.10	0.30	0.17
RGP_PRE	-	-	0.03	−0.07	0.05	−0.02	0.02
RGP_RHU	-	-	−0.04	−0.07	0.07	−0.02	0.00
RGP_DL	-	-	−0.06	0.14	0.08	−0.03	0.00
RGP_Rns	-	-	0.26	0.16	0.04	0.15	0.10

EPPA (number of effective panicles per unit area), TGPP (total number of grains per panicle), FGPP (filled grains per panicle), SSR (seed-setting rate), TGW (thousand-grain weight), PH (plant height), and PL (panicle length). EPPA and TGPP are determined before the heading stage and are not influenced by meteorological factors after the heading.

3.2. The Performance Comparison of Random Forest Model Based on Seven Different Feature Combinations

Figure 5 presents the performance of random forest yield prediction models for different input feature sets, using three metrics: root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2). To ensure reliable evaluation, the dataset is randomly split 100 times for each feature set, and the mean across the 100 iterations is reported as the final performance estimate.
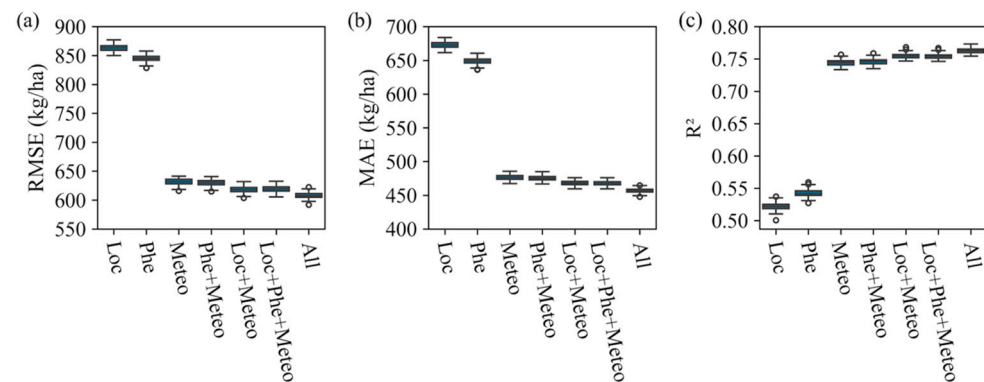
Figure 5. The average performance of the random forest regression model after 100 iterations based on different combinations of input variables. (**a**) Root mean square error (RMSE), (**b**) mean absolute error (MAE), and (**c**) coefficient of determination (R^2).

The RMSE (Figure 5a) and MAE (Figure 5b) reveal that models relying solely on geographical location (Loc) or phenological factors (Phen) generate large prediction errors, with RMSE > 845.10 kg/ha and MAE > 649.20 kg/ha. Relying exclusively on geographic location features (Loc), prediction error is particularly pronounced, with RMSE at 863.10 kg/ha and MAE at 672.75 kg/ha. Phen alone lowers the errors, but they remain high. Incorporating meteorological factors (Meteo) in the combined sets (Phen + Meteo, Loc + Meteo, and Loc + Phen + Meteo) markedly reduces both metrics, with RMSE values of 618.15–629.70 kg/ha and MAE values of 467.40–475.20 kg/ha. When all variables (All) are included—i.e., adding cultivar information to Loc, Phen, and Meteo—RMSE and MAE fall to 607.80 kg/ha and 456.75 kg/ha, respectively. Overall, integrating multiple features, especially Meteo, substantially lowers prediction error, and the inclusion of cultivar information provides further accuracy.

From the coefficient of determination (R^2, Figure 5c) perspective, different feature combinations show markedly different explanatory powers. Using only geographical variables (Loc), R^2 reaches 0.52, indicating weak independent predictive capacity. When only meteorological variables (Meteo) are used, R^2 significantly increases, approaching 0.75. With each addition of input variables, R^2 increases progressively. The combinations Loc + Meteo, Phen + Meteo, and Loc + Phen + Meteo all achieve $R^2 \geq 0.75$, providing robust predictions of rice yield variation. Including all variables (All) pushes R^2 to its maximum of 0.77. This shows that adding cultivar information to climate and phenological factors enhances overall accuracy and stability, yet cultivar data alone can lead to inaccurate yield estimates in large-scale predictions.

3.3. The Analysis of the Contribution of Feature Variables Based on the Optimal Model

Using the random forest model, rice yield is predicted from all features—geographic location (Loc), phenological factors (Phe), meteorological factors (Meteo), and cultivar characteristics (Cultivar). The model uses optimal parameters: maximum depth = 20, feature selection via the square root of the total number of features, minimum samples per leaf = 4, minimum samples to split = 10, 300 trees, and bootstrap = True. After 100 random splits, the best-performing run is displayed in Figure 6. On the training set, the model explains 89% of the variance, with RMSE = 407.39 kg/ha and MAE = 303.96 kg/ha. On the test set, it explains 77% of the variance, with RMSE = 601.45 kg/ha and MAE = 454.98 kg/ha.
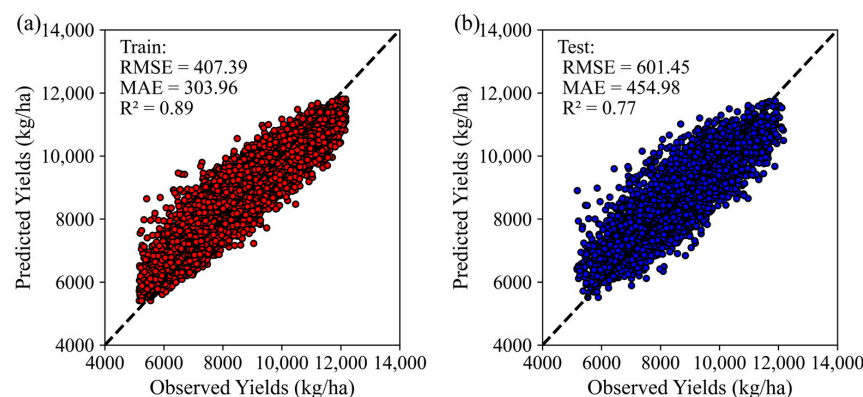
Figure 6. Model explainability of random forest regression model in train set (**a**) and test set (**b**).

Figure 7 presents the relative importance of yield-influencing features based on optimal parameters and the best full-factor model (the importance of each variable in all models can be referred to in the Table A3). Environmental variables account for the largest proportion (53.59%), followed by phenological variables (20.39%), geographic location variables (17.20%), and cultivar (8.82%). Among individual predictors, rice cultivar itself emerges as the single most critical factor. During VGP, daylength (VGP_DL, 8.07%) and net solar radiation (VGP_Rns, 6.69%) play significant roles in the process of yield formation. Additionally, key phenology events, including sowing date (DOY_Sow, 5.81%) and heading date (DOY_Hea, 6.55%), make substantial contributions to yield. Geographic characteristics such as elevation (Elev, 7.65%), latitude (Lat, 5.93%), and longitude (Lon, 3.64%) also demonstrate high importance. A closer look at meteorological variables across development periods reveals distinct requirements: during VGP, maximum temperature (VGP_TMax, 3.62%), accumulated precipitation (VGP_PRE, 3.30%), average temperature (VGP_TMean, 3.30%), and minimum temperature (VGP_TMin, 3.21%) demonstrate significant importance. In contrast, during RGP, minimum temperature (RGP_TMin, 3.09%) emerges as particularly important for yield prediction. These findings indicate that rice has different meteorological requirements during VGP and RGP. During the VGP, rice requires more sunlight and solar radiation, while during the RGP, yield is mainly affected by low temperatures. Overall, yield is governed by the integrated effects of cultivar, sowing date, meteorological conditions, and geographic factors. Cultivar exerts the strongest individual effect, while the sowing date interacts with meteorological variables to shape the final yield.
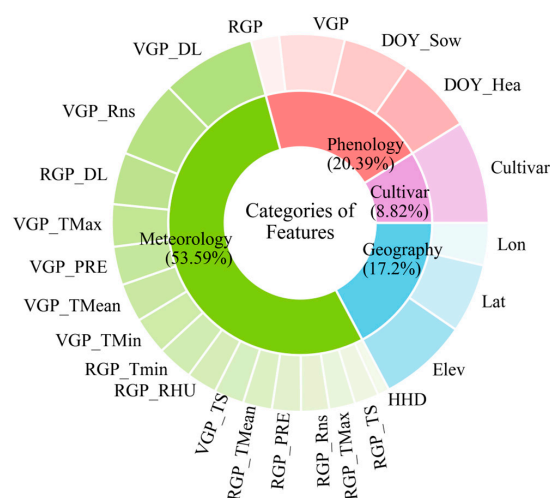

Figure 7. Relative importance of different input variables, including elevation (Elev), latitude (Lat), longitude (Lon), DOY_Sow (the day of year for sowing date), DOY_Hea (the day of year for heading

date), VGP (the vegetative growth period), RGP (the reproductive growth period), minimum temperature (TMin), maximum temperature (TMax), average temperature (TMean), accumulated precipitation (PRE), relative humidity (RHU), net solar radiation (Rns), daylength (DL), $\geq 8\ °C$ thermal summation (TS), and total days of high heat and high humidity (HHD).

Partial dependence plots reveal that multiple features affect rice yield through nonlinear mechanisms (Figure 8). Cultivar, Elev, VGP, RGP, VGP_Rns, and VGP_TS exhibit threshold effects, with stepwise yield shifts occurring at critical values. Latitude and longitude capture spatial yield variation across cultivation regions. Sowing date (DOY_Sow) and heading date (DOY_Hea) display pronounced seasonal influences, defining distinct optimal windows: the ideal sowing date falls between days 109 and 125 (mid-April to early May), and the optimal heading date spans days 189–246 (early July to early September). VGP_DL exerts a significant nonlinear positive effect; within the 14–14.7 h range, and each incremental increase in daylength steadily raises yield. Conversely, RGP_DL values above 13.7 h are associated with yield decline.
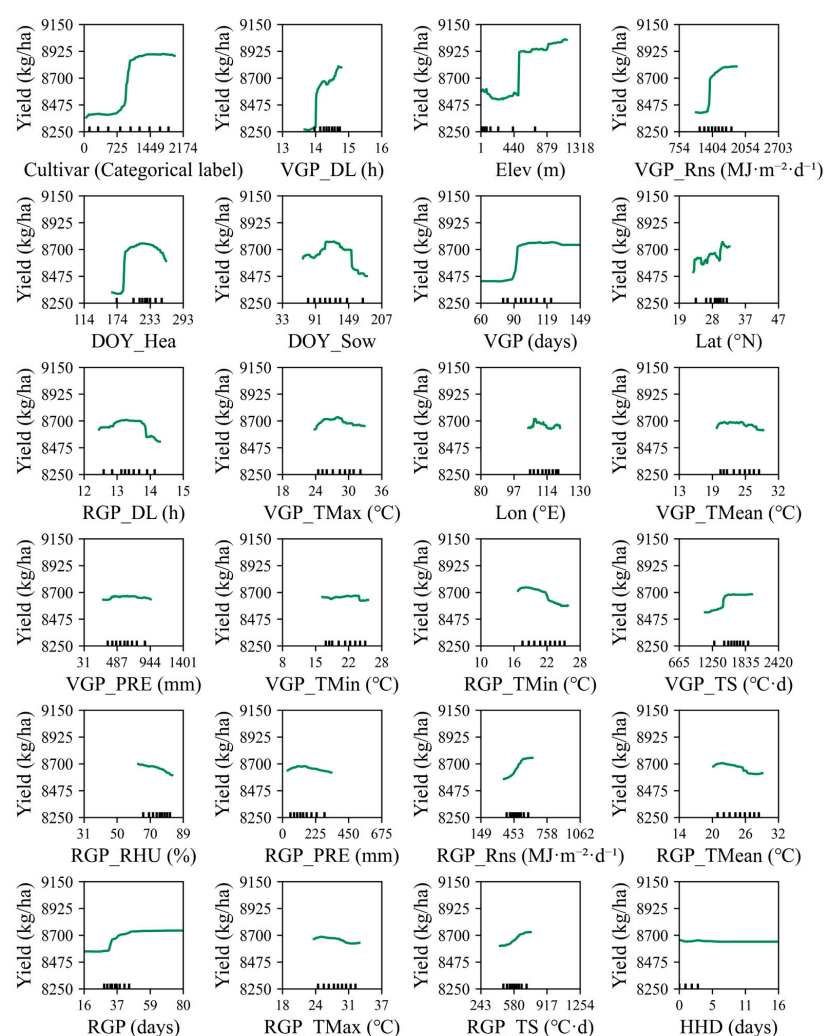

Figure 8. Partial dependence plots based on the importance of variables. Elev (elevation, m), Lat (latitude, °N), Lon (longitude, °E), DOY_Sow (the day of year for sowing date), DOY_Hea (the day of year for heading date), VGP (the vegetative growth period, days), RGP (the reproductive growth period, days), TMin (minimum temperature, °C), TMax (maximum temperature, °C), TMean (average temperature, °C), PRE (accumulated precipitation, mm), RHU (relative humidity, %), Rns (net solar radiation, $MJ \cdot m^{-2} \cdot d^{-1}$), DL (daylength, h), TS ($\geq 8\ °C$ thermal summation, °C d), and HHD (total days of high heat and high humidity, days). Cultivar represents the classification identifier and shows the impact on the yield. The major tick marks indicate coordinate values, while the minor tick marks represent data points.

4. Discussion

4.1. The Interaction Between Environment, Phenological Development, and Cultivar on Rice Yield

Based on multi-year trial data from China's rice-cultivation regions, this study reveals that rice yield exhibits significant differences across different cultivation site-years, mainly influenced by the interaction between geographic location (such as latitude and elevation), meteorological conditions, management measures (such as adjustment of sowing date), phenological development, and cultivar genetic traits.

Friedman's H-statistic was employed to quantify interaction strengths between variable groups in the random forest model, revealing complex interaction patterns among variable categories in rice yield formation. The cultivar–phenology interaction emerged as the most significant ($H^2 = 0.77$), demonstrating that rice varieties respond distinctively to different sowing dates and growth cycle durations. This finding emphasizes the importance of targeted variety breeding for specific planting seasons and growth cycle requirements. The cultivar–environment interaction followed in significance ($H^2 = 0.56$), highlighting variations in environmental adaptability across varieties and underscoring the necessity for regionalized variety distribution based on local environmental conditions [37]. Similarly important was the environment–phenology interaction ($H^2 = 0.53$), which illustrated how environmental factors substantially influence rice growth and developmental processes, suggesting that sowing dates and growth stage management should be tailored to local climate conditions [38]. The three-dimensional interaction among cultivar, environment, and phenology exhibited moderate strength ($H^2 = 0.25$). This represents the additional influence generated by the combined effect of all three factors beyond their two-dimensional interactions, indicating that incorporating these complex three-factor interactions can further enhance prediction accuracy in rice yield models.

The results of this study reveal complex spatial patterns across China. Latitude effects show yield increases from 19° N to peaks around 28–37° N, corresponding to China's rice cultivation zones from tropical and subtropical regions to the productive Middle–Lower Yangtze River areas. Longitude effects display yield fluctuations between 97 and 115° E and stability between 115 and 130° E, reflecting west-to-east transitions from mountainous southwestern regions to traditional rice-producing central provinces and eastern coastal areas. Wang et al. (2021) [39] also point out that due to different light–temperature conditions across the Yangtze River Basin, rice development period, crop economic traits, and yield from the upper reaches to the middle and lower reaches of the Yangtze River exhibit significant geographic variations. The apparent yield spike at 450 m elevation represents the optimal ecological niche formed at medium elevations (450–510 m) in the central mountainous transition zone, where moderate elevation combines with suitable latitude (27–32° N) to create ideal growing conditions. These non-monotonic patterns reflect China's complex geographic–ecological systems, where coordinates indirectly capture regional differences in climate, topography, rice varieties, and cultivation practices that collectively determine yield potential.

Consequently, geographic location not only encapsulates local climate information but also implies site-specific management strategies. The research results confirm that in high-latitude and high-elevation areas, adjusting the sowing date can improve climate conditions and regulate the process of development to enhance the yield. Because the higher the altitude, the worse the heat conditions will be, and the earlier the sowing date needs to be set to ensure the maturation [40]. Similarly, latitude influences grain weight and yield by altering temperature and solar radiation [41]. Regional environment variations will inevitably affect the genetic traits and crop yield. Among these factors, environmental conditions play a dominant role in driving yield variability, whereas the cultivar genotype establishes the potential yield ceiling; the actual yield outcome is ultimately governed by

the interaction between genotype and environment [42]. Thus, it is necessary to adopt adaptive management strategies across geographical gradients and climate zones. By employing environment similarity-based clustering approaches [43], agroecological zones with analogous climatic and geographical conditions are delineated. Subsequently, tailored adjustments to cultivation strategies—aligned with the environmental profiles of each distinct zone—are implemented, coupled with targeted cultivar selection (emphasizing stress tolerance and growth characteristics) to maximize regional production potential.

Phenological development also has a significant impact on rice yield. The research results indicate significant positive effects of VGP, RGP, and GP on yield, indicating that longer development periods favored dry matter accumulation and final yield formation. This finding is consistent with Wang et al. (2025) [44], who report that advancing sowing dates for single-season and early-season rice and delaying sowing for late-season rice extend their respective development periods and enhance potential yield. In this study, sowing date and heading date have shown weak linear correlations with yield, but exhibited nonlinear relationships, with heading date primarily driven by sowing date (Figures 4 and 8). These discoveries reveal a nonlinear saturation relationship between developmental duration and yield. Within the optimal sowing window, prolonging both vegetative and reproductive growth phases substantially enhances yield; conversely, sowing outside this window reduces the benefit through photo-thermal constraints. It should be noted that meteorological factors at different development periods have significant differences in their effects on yield. During VGP, net solar radiation, daylength, ≥ 8 °C thermal accumulation, and accumulated precipitation show positive correlations with yield, indicating that sunlight and temperature during this stage are critical for rice growth. In contrast, during RGP, daily minimum temperature, daily average temperature, and daily maximum temperature are significantly negatively correlated with yield. Especially, low temperature directly affects rice grain filling. These findings align with the views of Ma et al. (2025) [45], who demonstrate that low-temperature stress during grain filling markedly reduces seed-set rate, thousand-grain weight, and final yield, with the magnitude of decline intensifying as the duration of low-temperature stress increases.

The result of this study shows that rice cultivar is the most important factor affecting yield, with differences in its genetic characteristics playing a decisive role. In addition, genetic variations between rice cultivars can result in major differences in plant architecture, as research shows that even minor changes to the genetic makeup of rice can significantly change plant height, tillering characteristics, and productivity components [46]. This genetic basis for yield variation is clearly demonstrated in our findings, where cultivar emerged as the most important single factor influencing yield. The positive correlations between cultivar and panicle grain number and total grain number further indicate the potential influence of cultivar genetic background on yield. However, traits such as thousand-grain weight, seed setting rate, and number of effective panicles show relatively smaller relationships with yield, which may be related to the range of variation in these traits across different varieties and environments. Yield is determined collectively by the number of panicles per unit area, the number of grains per panicle, the seed-setting rate, and grain weight. Previous studies confirm that there is a significant compensation mechanism between the number of panicles per unit area and the number of grains per panicle, making it difficult to simultaneously increase both [47–49]. In the 1960s, high-density planting strategies were adopted in self-pollinated rice production to increase the number of panicles per unit area [50,51]. However, with the development of hybrid rice cultivars, the research direction gradually shifts toward increasing yield by enlarging the panicle size [52,53]. The findings suggest that the key approach for improving yield lies in adopting breeding

improvement, followed by controlling the number of effective panicles and increasing the total grain count per panicle.

4.2. The Key Factors of Modeling Rice Yield

The random forest algorithm demonstrates strong performance in analyzing the combined effects of genetic traits, environmental factors, and management practices on crop development and yield formation [54,55]. By constructing random forest regression models with multiple input variable combinations, this study systematically evaluates the contributions of geographic location (Loc), phenology (Phe), meteorology (Meteo), and rice cultivar (Cultivar) to modeling rice yield.

Variable importance ranking based on all variables model identified the top factors for rice yield prediction: rice cultivar (cultivar, 8.82%), average daylength during vegetative growth period (VGP_DL, 8.07%), elevation (Elev, 7.65%), accumulated radiation during the vegetative growth period (VGP_Rns, 6.69%), the day of year for heading date (DOY_Hea, 6.55%), latitude (Lat, 5.93%), the day of year for sowing date (DOY_Sow, 5.81%), and vegetative growth period (VGP, 5.61%). These account for over 50% of prediction importance. Among them, rice cultivar emerges as the most influential single factor, highlighting its fundamental role in genetic control [19]. This finding is consistent with Cheng (2021), who reports that China's rice breeding achievements have advanced national average rice yields from 3000 kg/ha in the 1950s to 7050 kg/ha currently, representing a 2.35-fold increase [56]. The relatively high importance of site geographic location (longitude, latitude, and elevation) aligns with findings by Xu et al. (2025) [57] and Wang et al. (2024) [58].

Meteorological factors closely associated with rice phenological development play decisive roles in yield formation [59]. Among meteorological variables, light conditions (VGP_DL, VGP_Rns, and RGP_DL) are most critical for rice yield prediction. While reproductive growth period daylength (RGP_DL) is not traditionally considered a direct yield determinant, its importance emerges through several mechanisms. RGP_DL indirectly reflects the seasonal position of heading date (related to DOY_Hea), captures seasonal temperature effects on pollen viability and grain filling when combined with reproductive growth period temperature (RGP_TMin), and potentially indicates compatibility between cultivar and heading date adjustment, demonstrating genotype–environment interactions. These findings align with those of Zhao et al. (2017) [60] and Li et al. (2023) [61], who report that substantial temporal variability in the importance of meteorological factors. Although heat–humidity days (HHD) currently show the lowest importance among input variables in our dataset, this is primarily due to their infrequent occurrence in the present dataset rather than an actual lack of biological relevance. In fact, HHD creates favorable conditions for major rice diseases, such as blast disease, which can destroy enough rice annually to feed 60 million people [62]. Given that global warming is expected to significantly increase the frequency of high-temperature and high-humidity events [63], HHD should be retained as a critical input variable in predictive models.

The partial dependence plots reveal the complex interactions between crop growth and environmental factors. Among them, cultivar exhibits a dominant influence on yield, highlighting the crucial role of genetic improvement through breeding. Elevation shows a clear gradient response, while latitude and longitude exhibit pronounced fluctuations, reflecting the spatial heterogeneity in climatic suitability. Both the sowing date (DOY_Sow) and flowering date display evident optimal windows, underscoring the value of precise selection of planting periods. The vegetative growth period (VGP) and reproductive growth period (RGP) exhibit plateau-like relationships with yield, suggesting threshold effects in phenological duration—both excessively short durations and long durations negatively impact yield formation. Temperature-related variables, including VGP_Tmax, VGP_Tmin,

VGP_TS, RGP_TMin, and RGP_TMean, demonstrate marginal effects, highlighting the presence of critical temperature thresholds and the necessity for coordinated optimization in cultivation management. Overall, the random forest model effectively captures nonlinear and interactive effects among multiple agroecological drivers, offering theoretical insights and practical guidance for climate-resilient rice production strategies.

Variable importance analysis results show that cultivar type, growth period length, radiation, day length, elevation, and latitude each contribute less than 10% individually to yield prediction, yet they remain key variables due to the complex and multifactorial nature of crop productivity. This apparent paradox reflects the multidimensional, nonlinear coupling that drives yield formation. In ensemble models like random forest, explanatory power becomes distributed across numerous correlated but complementary variables, preventing any single factor from dominating. Multicollinearity among spatial–temporal variables (latitude, day length, radiation, and temperature) further disperses importance metrics among related feature clusters. Mechanistically, cultivars (genotype, G) determine potential yield ceilings and source–sink relationships, while development period lengths (VGP/RGP/GP) serve as integrated indicators of environment–management interactions that influence biomass accumulation and grain filling duration. Net radiation and day length control photosynthetic quantum supply and photoperiodic signaling, directly affecting canopy energy absorption and phenological transitions. Altitude and latitude function as spatial proxy variables that capture temperature gradients, radiation fields, photoperiod trajectories, and precipitation patterns. The remaining variance is explained by equally important but diffuse factors, including temperature thresholds, precipitation, heat–moisture stress, and interannual climate variability, which alternately become limiting factors across different ecological zones and growing seasons. This distributed importance pattern actually reflects the high-dimensional complexity of yield formation processes rather than diminishing the significance of identified key drivers.

4.3. Limitations and Prospects

The model developed in this study relies on data from agricultural experiment sites using optimal management practices that differ significantly from actual farming conditions. The cultivar trial data fail to fully capture the complexity and variability of real agricultural environments, particularly in northern ecological zones, limiting model generalizability in regions with extreme climates or lower management levels. Additionally, this study omits variables such as soil texture, organic matter, and nutrient content, which may significantly influence yield and interact complexly with other factors. While location-related effects partly reflect soil differences, future research requires systematic soil data collection. Technological advancement during the study period (2007–2018) is not explicitly considered, despite China's rice production benefiting from improvements in genetics, crop protection, and agricultural machinery that collectively contribute to yield growth. Furthermore, the model only accounts for interannual climate variability without comprehensively evaluating long-term climate change impacts, as future climatic conditions will increasingly diverge from historical patterns, challenging historical-based models in predicting extreme weather events and resulting pest and disease dynamics.

Future research will integrate phenological process models, climate change simulations, and yield prediction frameworks, extending the scope to grain quality, pest and disease risks, and climate adaptation strategies. Systematic collection of soil data and regional technology adoption rates will help distinguish between genetic improvement, agronomic advancement, and their interactions with environmental and phenological factors. This integrated approach will support the development of comprehensive agricultural modeling systems adapted to changing environments, providing scientific evidence for

policy formulation and field management. Moreover, future research will also aim to investigate genetic–environment interactions by modeling the correlation between specific genetic markers in rice cultivars and their performance across diverse cultivation environments. This approach would involve genotyping the studied rice cultivars and correlating genetic information with phenotypic performance under varying environmental conditions. Such research could identify key genetic determinants of environmental adaptability and yield stability, ultimately supporting more targeted breeding programs for climate resilience.

5. Conclusions

Based on observational data from rice cultivar trials (2007–2018), this study systematically examined cultivar–environment–phenology interactions through random forest modeling, revealing distinct yield response patterns across multiple factors. The cultivar exhibited strong positive threshold effects, with significant yield jumps for specific genotypes. Elevation showed step-like positive nonlinear relationships with notable yield increases at approximately 450 m, while latitude and longitude displayed fluctuating nonlinear patterns that collectively captured the intricate geographical ecosystem of China's rice-growing regions, defining specific climatic conditions and ecological niches. Sowing and heading dates showed inverted U-shaped relationships, identifying mid-April to early May as the optimal sowing window for yield maximization. VGP, VGP_DL, VGP_Rns, RGP_TS, and RGP demonstrated positive threshold effects, while RGP_DL showed a negative threshold effect. Temperature parameters during VGP (including VGP_TMax, VGP_TMean, and VGP_TMin), VGP_PRE, RGP_TMax, and RGP_PRE had relatively gentle effects on yield variation. RGP_TMin, RGP_TMean, and RGP_RHU had negative effects on yield. The impact of total days of high heat and high humidity (HHD) was minimal. This study provided a robust framework for understanding rice yield formation mechanisms and offered practical guidance for cultivar selection and agronomic management optimization across diverse agroecological environments.

Abbreviations

The following abbreviations are used in this manuscript:

Lat	Latitude.
Lon	Longitude.
Elev	Elevation.
DOY_Sow	The day of year for sowing date.
DOY_Hea	The day of year for heading date.
DOY_Mat	The day of year for maturation date.
GP	The entire development period from sowing to maturation.
VGP	The vegetative growth period from sowing to heading.
RGP	The reproductive growth period from heading to maturation.
EPPA	The number of effective panicles per unit area.
TGPP	Total number of grains per panicle.
FGPP	Filled grains per panicle.
SSR	Seed-setting rate.
TGW	Thousand-grain weight.
PH	Plant height.
PL	Panicle length.
TMin	Minimum temperature.
TMax	Maximum temperature.
TMean	Mean temperature.
PRE	Accumulated precipitation.
RHU	Relative humidity.
Rns	Net solar radiation.
TS	$\geq 8\,°C$ thermal summation.
DL	Daylength.
HN	The frequency of high-temperature events (3 consecutive days with daily average temperature $\geq 30\,°C$).
HD	Total days of HN.
HDD	Accumulated heat of HD.
CN	The frequency of cold damage events (daily mean temperature $\leq 17\,°C$ above $36°$ N or $\leq 20\,°C$ below $36°$ N for 3 consecutive days).
CD	Total days of CN.
CDD	Accumulated cold of CD.
HHD	Total days with high-heat and high-humidity (daily average temperature $\geq 25.0\,°C$ and relative humidity $\geq 90.0\%$).

Appendix A

Appendix A.1

Table A1 lists all variables used in this study, their abbreviations, descriptions, and whether they were used as input variables in the seven models developed.

Table A1. All variables, abbreviations, descriptions, units, and their applications in the seven models developed in this study.

Category	Variable	Abbreviation	Description	Unit	Model
Geographic variable **(Loc)**	Longitude	Lon	Spatial coordinate (east–west)	$^\circ$ E	1, 4, 6, 7
	Latitude	Lat	Spatial coordinate (north–south)	$^\circ$ N	1, 4, 6, 7
	Elevation	Elev	Height above sea level	m	1, 4, 6, 7
Phenological variable **(Phen)**	Sowing date	DOY_Sow	Day of year for sowing	-	2, 5, 6, 7
	Heading date	DOY_Hea	Day of year for heading	-	2, 5, 6, 7
	Vegetative growth period	VGP	Days from sowing to heading	Days	2, 5, 6, 7
	Reproductive growth period	RGP	Days from heading to maturity	Days	2, 5, 6, 7
	The entire development period	GP	Days from sowing to maturity	Days	-
Meteorological variable **(Meteo)**	During Vegetative Growth Period (VGP)				
	Mean minimum temperature	VGP_TMin	Average daily minimum temperature during VGP	$^\circ$C	3, 4, 5, 6, 7
	Mean maximum temperature	VGP_TMax	Average daily maximum temperature during VGP	$^\circ$C	3, 4, 5, 6, 7
	Mean temperature	VGP_TMean	Average daily mean temperature during VGP	$^\circ$C	3, 4, 5, 6, 7
	Accumulated precipitation	VGP_PRE	Total precipitation during VGP	mm	3, 4, 5, 6, 7
	Average relative humidity	VGP_RHU	Average daily relative humidity during VGP	%	-
	Accumulated net solar radiation	VGP_Rns	Total net solar radiation during VGP	$MJ \cdot m^{-2} \cdot d^{-1}$	3, 4, 5, 6, 7
	Average daylength	VGP_DL	Average daily photoperiod during VGP	h	3, 4, 5, 6, 7
	Thermal summation	VGP_TS	$\geq 8\,^\circ$C thermal summation during VGP	$^\circ$C·d	3, 4, 5, 6, 7
	During Reproductive Growth Period (RGP)				
	Mean minimum temperature	RGP_TMin	Average daily minimum temperature during RGP	$^\circ$C	3, 4, 5, 6, 7
	Mean maximum temperature	RGP_TMax	Average daily maximum temperature during RGP	$^\circ$C	3, 4, 5, 6, 7
	Mean temperature	RGP_TMean	Average daily mean temperature during RGP	$^\circ$C	3, 4, 5, 6, 7
	Accumulated precipitation	RGP_PRE	Total precipitation during RGP	mm	3, 4, 5, 6, 7
	Average relative humidity	RGP_RHU	Average daily relative humidity during RGP	%	3, 4, 5, 6, 7
	Accumulated net solar radiation	RGP_Rns	Total net solar radiation during RGP	$MJ \cdot m^{-2} \cdot d^{-1}$	3, 4, 5, 6, 7

Table A1. *Cont.*

Category	Variable	Abbreviation	Description	Unit	Model
	During Vegetative Growth Period (VGP)				
	Average daylength	RGP_DL	Average daily photoperiod during RGP	h	3, 4, 5, 6, 7
	Thermal summation	RGP_TS	≥ 8 °C thermal summation during RGP	°C·d	3, 4, 5, 6, 7
	During the Entire Growth Period (GP)				
	Mean minimum temperature	GP_TMin	Average daily minimum temperature during GP	°C	-
	Mean maximum temperature	GP_TMax	Average daily maximum temperature during GP	°C	-
	Mean temperature	GP_TMean	Average daily mean temperature during GP	°C	-
	Accumulated precipitation	GP_PRE	Total precipitation during GP	mm	-
	Average relative humidity	GP_RHU	Average daily relative humidity during GP	%	-
	Accumulated net solar radiation	GP_Rns	Total net solar radiation during GP	$MJ \cdot m^{-2} \cdot d^{-1}$	-
	Average daylength	GP_DL	Average daily photoperiod during GP	h	-
	Thermal summation	GP_TS	≥ 8 °C thermal summation during GP	°C·d	-
	Stress Indicators				
Meteorological variable (**Meteo**)	The frequency of high-temperature events	HN	A high temperature event refers to a period when the daily average temperature is ≥ 30 °C for three consecutive days	-	-
	Total days of HN	HD	Total days of HN during GP	Days	-
	Accumulated heat of HD	HDD	≥ 30 °C thermal summation during high temperature events throughout the entire growing period	°C·d	-
	The frequency of cold damage events	CN	A cold damage event refers to a period when the daily average temperature is ≤ 17 °C above 36° N or ≤ 20 °C below 36° N for consecutive 3 days	-	-
	Total days of CN	CD	Total days of CN during GP	Days	-
	Accumulated cold of CD	CDD	Accumulated chilling (≤ 17 °C above 36° N or ≤ 20 °C below 36° N) during cold damage events throughout the entire growing period	°C·d	-
	High-heat and high-humidity days	HHD	Days with mean temperature ≥ 25 °C and relative humidity $\geq 90\%$	Days	3, 4, 5, 6, 7

Table A1. *Cont.*

Category	Variable	Abbreviation	Description	Unit	Model
	Rice cultivar	**Cultivar**	Unique identifier for each rice cultivar (Categorical label)	-	7
	The number of effective panicles per unit area	EPPA	-	10^4 panicles/ha	-
	Total number of grains per panicle	TGPP	-	gains/panicle	-
Cultivar variables	Filled grains per panicle	FGPP	-	gains/panicle	-
	Seed-setting rate	SSR	-	%	-
	Thousand-grain weight	TGW	-	g	-
	Plant height	PH	-	cm	-
	Panicle length	PL	-	cm	-

Model: 1. Geographic variables only (Loc); 2. Phenological variables only (Phen); 3. Meteorological variables only (Meteo); 4. Geographic + Meteorological variables (Loc + Meteo); 5. Phenological + Meteorological variables (Phen + Meteo); 6. Geographical + Phenological + Meteorological variables (Loc + Phen + Meteo); 7. All variables (Loc + Phen + Meteo + Cultivar). Note: The rice cultivar was treated as a categorical variable, with each cultivar encoded with a unique identifier. This approach allowed the model to capture cultivar-specific effects without imposing any numerical ordering or relationship between different cultivars.

Appendix A.2

Table A2 provides the statistical distribution for all variables incorporated in the model development during the experiments.

Table A2. Statistical distribution for all variables incorporated in the model development.

Variable	Min	25%	50%	Mean	75%	Max	Standard Deviation	Coefficient of Variation
Yield (kg/ha)	5162.55	7792.50	8668.58	8664.00	9544.50	12,178.80	1290.00	0.15
Cultivar (Categorical label)	0	-	-	-	-	2174	-	-
Lat (° N)	19.15	27.25	29.43	29.02	31.02	46.67	3.31	0.11
Lon (° E)	80.12	107.73	113.12	112.70	117.23	130.50	5.39	0.05
Elev (m)	1.00	28.20	75.60	237.43	325.00	1318.00	336.77	1.42
DOY_Sow	33	92	115	119.6	137	207	34.7	0.3
DOY_Hea	114	209	224	220.5	237	293	28.3	0.1
VGP (days)	60	87	100	100.9	114	149	16.0	0.2
RGP (days)	16	32	36	36.7	41	80	6.8	0.2
VGP_TMean (°C)	12.93	21.60	24.36	24.27	26.72	31.75	2.90	0.12
VGP_TMax (°C)	18.44	25.77	28.42	28.24	30.46	35.73	2.83	0.10
VGP_TMin (°C)	8.25	17.97	20.95	20.80	23.44	28.35	3.13	0.15
VGP_TS (°C·d)	664.65	1499.75	1637.61	1613.38	1767.54	2419.91	235.19	0.15
VGP_PRE (mm)	30.55	447.87	588.62	598.02	722.02	1400.55	200.93	0.34
VGP_DL (h)	12.61	14.04	14.30	14.26	14.54	16.20	0.41	0.03
VGP_Rns ($MJ·m^{-2}·d^{-1}$)	754.07	1287.31	1461.71	1466.88	1634.70	2703.49	241.41	0.16
RGP_TMean (°C)	14.13	22.51	24.88	24.73	27.01	31.72	2.79	0.11
RGP_TMax (°C)	17.60	26.40	28.71	28.59	30.79	37.49	2.86	0.10

Table A2. *Cont.*

Variable	Min	25%	50%	Mean	75%	Max	Standard Deviation	Coefficient of Variation
RGP_TMin (°C)	10.33	19.20	21.63	21.43	23.73	27.89	2.83	0.13
RGP_TS (°C·d)	243.27	524.59	583.12	586.77	642.87	1253.94	95.20	0.16
RGP_PRE (mm)	0.04	89.08	139.90	158.55	214.19	674.96	93.93	0.59
RGP_RHU (%)	30.63	70.16	75.22	74.20	78.94	89.19	6.31	0.08
RGP_DL (h)	11.83	13.08	13.46	13.48	13.97	15.44	0.66	0.05
RGP_Rns ($MJ \cdot m^{-2} \cdot d^{-1}$)	148.80	427.52	476.19	482.44	529.85	1062.14	82.34	0.17
HHD (days)	0	0	0	1.1	2	16	1.6	1.5

Appendix A.3

Table A3 provides the input variable categories and their importance in each random forest regression model used in this study.

Table A3. The importance of individual variables in each model.

Variable Importance (%)	Variable Categories Input into the Model						
	All	Loc	Phe	Meteo	Phe + Meteo	Loc + Meteo	Loc + Phe + Meteo
Cultivar	8.82						
VGP_DL	8.07			14.35	9.59	11.40	8.64
Elev	7.65	32.37				9.87	7.85
VGP_Rns	6.69			12.17	8.72	9.46	7.46
DOY_Hea	6.55		27.96		8.18		6.36
Lat	5.93	43.72				7.99	6.06
DOY_Sow	5.81		28.51		8.08		6.16
VGP	5.61		24.54		6.81		6.30
RGP_DL	4.44			10.05	6.01	6.88	4.78
Lon	3.64	23.91				4.42	3.97
VGP_TMax	3.62			6.54	5.06	5.19	4.29
VGP_PRE	3.30			5.90	4.83	4.72	3.95
VGP_TMean	3.30			6.40	4.72	4.83	3.79
VGP_TMin	3.21			6.42	4.67	4.66	3.65
RGP_TMin	3.09			5.69	4.70	4.09	3.51
RGP_RHU	2.60			4.49	3.85	3.72	3.20
VGP_TS	2.57			5.26	3.06	4.52	2.62
RGP_TMean	2.49			4.50	3.72	3.39	2.85
RGP_PRE	2.48			4.32	3.62	3.59	2.97
RGP	2.42		18.98		3.53		2.78
RGP_Rns	2.41			4.81	3.48	3.63	2.71
RGP_TMax	2.23			3.73	3.18	3.10	2.54
RGP_TS	2.09			3.55	2.62	3.27	2.34
HHD	1.00			1.82	1.58	1.28	1.22

References

1. Global Food and Agriculture Statistics of FAO 2024. Available online: https://openknowledge.fao.org/server/api/core/bitstreams/d784864f-7f28-49d2-903e-6680d09a9d97/content/cd2971en.html (accessed on 19 September 2025).

2. Bergeret, P. *The Future of Food and Agriculture: Trends and Challenges*, 1st ed.; Food and Agriculture Organization of the United Nations: Rome, Italia, 2017; pp. 39–56.

3. Shukla, P.R.; Skeg, J.; Buendia, E.C.; Masson-Delmotte, V.; Pörtner, H.O.; Roberts, D.C.; Zhai, P.; Slade, S.R.; Connors, S.; van Diemen, S.; et al. *Climate Change and Land: An IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable*

Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems, 1st ed.; Intergovernmental Panel on Climate Change: Geneva, Switzerland, 2020; pp. 437–550.

4. Zhao, Z.; Wang, E.; Kirkegaard, J.A.; Rebetzke, G.J. Novel Wheat Varieties Facilitate Deep Sowing to Beat the Heat of Changing Climates. *Nat. Clim. Change* **2022**, *12*, 291–296. [CrossRef]

5. Qian, Q.; Guo, L.; Smith, S.M.; Li, J. Breeding High-Yield Superior Quality Hybrid Super Rice by Rational Design. *Natl. Sci. Rev.* **2016**, *3*, 283–294. [CrossRef]

6. Lu, M. Impact of Climate Change on Rice and Adaptation Strategies: A Review. *Adv. Res. Res.* **2024**, *4*, 252–262. [CrossRef]

7. Pan, Y.; Su, Z. A review on the growth simulation of rice during its growth period. *J. Zhejiang Univ. Sci.* **2011**, *2*, 434–438.

8. Atkinson, D.; Porter, J.R. Temperature, Plant Development and Crop Yields. *Trends Plant Sci.* **1996**, *1*, 119–124. [CrossRef]

9. Bloomfield, M.T.; Celestina, C.; Hunt, J.R.; Huth, N.; Zheng, B.; Brown, H.; Zhao, Z.; Wang, E.; Stefanova, K.; Hyles, J.; et al. Vernalisation and Photoperiod Responses of Diverse Wheat Genotypes. *Crop Pasture Sci.* **2023**, *74*, 405–422. [CrossRef]

10. Liu, S.; Baret, F.; Abichou, M.; Manceau, L.; Andrieu, B.; Weiss, M.; Martre, P. Importance of the Description of Light Interception in Crop Growth Models. *Plant Physiol.* **2021**, *186*, 977–997. [CrossRef]

11. Tao, F.; Hayashi, Y.; Zhang, Z.; Sakamoto, T.; Yokozawa, M. Global Warming, Rice Production, and Water Use in China: Developing a Probabilistic Assessment. *Agric. For. Meteorol.* **2008**, *148*, 94–110. [CrossRef]

12. World Bank Group Water in Agriculture: Towards Sustainable Agriculture. Available online: http://documents.worldbank.org/curated/en/875921614166983369 (accessed on 19 September 2025).

13. Vergara, B.S.; Chang, T.T. *The Flowering Response of the Rice Plant to Photoperiod: A Review of the Literature*, 4th ed.; International Rice Research Institute: Manila, Philippines, 1985; pp. 1–26.

14. Yin, X.; Kropff, M.J.; Horie, T.; Nakagawa, H.; Centeno, H.G.S.; Zhu, D.; Goudriaan, J. A Model for Photothermal Responses of Flowering in Rice I. Model Description and Parameterization. *Field Crop Res.* **1997**, *51*, 189–200. [CrossRef]

15. Awan, M.I.; van Oort, P.A.J.; Bastiaans, L.; van der Putten, P.E.L.; Yin, X.; Meinke, H. A Two-Step Approach to Quantify Photothermal Effects on Pre-Flowering Rice Phenology. *Field Crop Res.* **2014**, *155*, 14–22. [CrossRef]

16. Shimono, H.; Okada, M.; Kanda, E.; Arakawa, I. Low Temperature-Induced Sterility in Rice: Evidence for the Effects of Temperature before Panicle Initiation. *Field Crop Res.* **2007**, *101*, 221–231. [CrossRef]

17. Jagadish, S.; Craufurd, P.; Wheeler, T. High Temperature Stress and Spikelet Fertility in Rice (*Oryza sativa* L.). *J. Exp. Bot.* **2007**, *58*, 1627–1635. [CrossRef] [PubMed]

18. Xiong, D.; Ling, X.; Huang, J.; Peng, S. Meta-Analysis and Dose-Response Analysis of High Temperature Effects on Rice Yield and Quality. *Environ. Exp. Bot.* **2017**, *141*, 1–9. [CrossRef]

19. Zainuddin, F.; Ismail, M.R.; Hatta, M.A.M.; Ramlee, S.I. Advancement in Modern Breeding and Genomic Approaches to Accelerate Rice Improvement: Speed Breeding Focus. *Euphytica* **2024**, *220*, 109. [CrossRef]

20. Jia, Z. Study on photothermal reaction characteristics of Shanyou63. *Cult. Plant.* **1989**, *16*, 49–53. [CrossRef]

21. Liang, G.; Cai, S. Study on the Critical Day Length of Panicle Emergence in Rice Varieties. *J. South China Agric. Univ.* **1980**, *1*, 54–66.

22. Summerfield, R.J.; Collinson, S.T.; Ellis, R.H.; Roberts, E.H.; De Vries, F.W.T.P. Photothermal Responses of Flowering in Rice (*Oryza sativa*). *Ann. Bot.* **1992**, *69*, 101–112. [CrossRef]

23. Nandram, B.; Berg, E.; Barboza, W. A Hierarchical Bayesian Model for Forecasting State-Level Corn Yield. *Environ. Ecol. Stat.* **2014**, *21*, 507–530. [CrossRef]

24. Lobell, D.B.; Asseng, S. Comparing Estimates of Climate Change Impacts from Proce-ssbased and Statistical Crop Models. *Environ. Res. Lett.* **2017**, *12*, 015001. [CrossRef]

25. Mathieu, J.A.; Aires, F. Assessment of the Agro-Climatic Indices to Improve Crop Yield Forecasting. *Agric. For. Meteorol.* **2018**, *253*, 15–30. [CrossRef]

26. Sulṭānī, A.; Sinclair, T.R. *Modeling Physiology of Crop Development, Growth and Yield*; CABI: London, UK, 2012; pp. 154–196.

27. Jones, J.W.; Hoogenboom, G.; Porter, C.H.; Boote, K.J.; Batchelor, W.D.; Hunt, L.A.; Wilkens, P.W.; Singh, U.; Gijsman, A.J.; Ritchie, J.T. The DSSAT Cropping System Model. *Eur. J. Agron.* **2003**, *18*, 235–265. [CrossRef]

28. Basso, B.; Liu, L.; Ritchie, J.T. A Comprehensive Review of the CERES-Wheat, -Maize and -Rice Models' Performances. *Adv. Agron.* **2016**, *136*, 27–132.

29. 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]

30. Bouman, B.A.M.; Kropff, M.J.; Tuong, T.P.; Wopereis, M.C.S.; ten Berge, H.F.M.; van Laar, H.H. *ORYZA2000: Modeling Lowland Rice*; IRRI: Manila, Philippines, 2001; pp. 25–76.

31. Zhao, M.; Peng, C.; Xiang, W.; Deng, X.; Tian, D.; Zhou, X.; Yu, G.; He, H.; Zhao, Z. Plant Phenological Modeling and Its Application in Global Climate Change Research: Overview and Future Challenges. *Environ. Rev.* **2013**, *21*, 1–14. [CrossRef]

32. Silva, J.V.; Giller, K.E. Grand Challenges for the 21st Century: What Crop Models Can and Can't (yet) Do. *J. Agric. Sci.* **2020**, *158*, 794–805. [CrossRef]

33. Betts, R.A. Integrated Approaches to Climate–Crop Modelling: Needs and Challenges. *Philos. Trans. R. Soc. B Int. J. Biol. Sci.* **2005**, *360*, 2049–2065. [CrossRef]

34. Challinor, A.J.; Ewert, F.; Arnold, S.; Simelton, E.; Fraser, E. Crops and Climate Change: Progress, Trends, and Challenges in Simulating Impacts and Informing Adaptation. *J. Exp. Bot.* **2009**, *60*, 2775–2789. [CrossRef]

35. Yang, S.; Hu, X. *China Rice New Cultivar Trials (2007–2018)*; China Agricultural Science and Technology Press: Beijing, China, 2019.

36. Keisling, T.C. Calculation of the Length of Day. *Agron. J.* **1982**, *74*, 758–759. [CrossRef]

37. Parihar, A.K.; Hazra, K.K.; Lamichaney, A.; Gupta, D.S.; Kumar, J.; Mishra, R.K.; Singh, A.K.; Bhartiya, A.; Sofi, P.A.; Lone, A.A.; et al. Multi-Location Evaluation of Field Pea in Indian Climates: Eco-Phenological Dynamics, Crop-Environment Relationships, and Identification of Mega-Environments. *Int. J. Biometeorol.* **2024**, *68*, 1973–1987. [CrossRef] [PubMed]

38. Brinkhoff, J.; McGavin, S.L.; Dunn, T.; Dunn, B.W. Predicting Rice Phenology and Optimal Sowing Dates in Temperate Regions Using Machine Learning. *Agron. J.* **2024**, *116*, 871–885. [CrossRef]

39. Wang, D.; Li, X.; Ye, C.; Xu, C.; Chen, S.; Chu, G.; Zhang, Y.; Zhang, X. Geographic Variation in the Yield Formation of Single-Season High-Yielding Hybrid Rice in Southern China. *J. Integr. Agric.* **2021**, *20*, 438–449. [CrossRef]

40. Wang, C.; Zhang, Z.; Zhang, J.; Tao, F.; Chen, Y.; Ding, H. The Effect of Terrain Factors on Rice Production: A Case Study in Hunan Province. *J. Geogr. Sci.* **2019**, *29*, 287–305. [CrossRef]

41. Baloch, N.; Liu, W.; Hou, P.; Ming, B.; Xie, R.; Wang, K.; Liu, Y.; Li, S. Effect of Latitude on Maize Kernel Weight and Grain Yield across China. *Agron. J.* **2021**, *113*, 1172–1182. [CrossRef]

42. Bustos-Korts, D.; Malosetti, M.; Chenu, K.; Boer, P.M.; Chapman, S.; Zheng, B.; van Eeuwijk, A.F. From QTLs to Adaptation Landscapes: Using Genotype-To-Phenotype Models to Characterize G×E Over Time. *Front. Plant Sci.* **2019**, *10*, 1540. [CrossRef]

43. Tan, J.; Zhao, G.; Tian, Q.; Zheng, L.; Kang, X.; He, Q.; Shi, Y.; Chen, B.; Wu, D.; Yao, N.; et al. Overcoming Mechanistic Limitations of Process-Based Phenological Models: A Data Clustering Method for Large-Scale Applications. *Agric. For. Meteorol.* **2024**, *356*, 110167. [CrossRef]

44. Wang, Y.; Li, S.; Zhao, J.; Wang, C.; Feng, Y.; Zhao, M.; Shi, X.; Chen, F.; Chu, Q. Planting Date Adjustment and Varietal Replacement Can Effectively Adapt to Climate Warming in China Southern Rice Area. *Agr. Syst.* **2025**, *226*, 104330. [CrossRef]

45. Ma, H.; Jia, Y.; Wang, W.; Wang, J.; Zou, D.; Wang, J.; Gong, W.; Han, Y.; Dang, Y.; Wang, J.; et al. Effects of Low-Temperature Stress During the Grain-Filling Stage on Carbon–Nitrogen Metabolism and Grain Yield Formation in Rice. *Agron. J.* **2025**, *15*, 417. [CrossRef]

46. Adnan, M.R.; Wilujeng, E.D.I.; Aisyah, M.D.N.; Alif, T.; Galushasti, A. Preliminary Agronomic Characterization of Japonica Rice Carrying a Tiller Number Mutation under Greenhouse Conditions. *Dysona Appl. Sci.* **2025**, *6*, 291–299. [CrossRef]

47. Miller, B.C.; Hill, J.E.; Roberts, S.R. Plant Population Effects on Growth and Yield in Water-Seeded Rice. *Agron. J.* **1991**, *83*, 291–297. [CrossRef]

48. Gravois, K.A.; Helms, R.S. Path Analysis of Rice Yield and Yield Components as Affected by Seeding Rate. *Agron. J.* **1992**, *84*, 1–4. [CrossRef]

49. Ottis, B.V.; Talbert, R.E. Rice Yield Components as Affected by Cultivar and Seeding Rate. *Agron. J.* **2005**, *97*, 1622–1625. [CrossRef]

50. Fagade, S.O.; De Datta, S.K. Leaf Area Index, Tillering Capacity, and Grain Yield of Tropical Rice as Affected by Plant Density and Nitrogen Level1. *Agron. J.* **1971**, *63*, 503–506. [CrossRef]

51. Wu, G.; Wilson, L.T.; McClung, A.M. Contribution of Rice Tillers to Dry Matter Accumulation and Yield. *Agron. J.* **1998**, *90*, 317–323. [CrossRef]

52. Ma, J.; Ma, W.; Ming, D.; Yang, S.; Zhu, Q. Characteristics of Rice Plant with Heavy Panicle. *Agric. Sci. China* **2006**, *5*, 911–918. [CrossRef]

53. Meng, T.; Wei, H.; Li, C.; Dai, Q.; Xu, K.; Huo, Z.; Wei, H.; Guo, B.; Zhnag, H. Morphological and Physiological Traits of Large-Panicle Rice Varieties with High Filled-Grain Percentage. *J. Integr. Agric.* **2016**, *15*, 1751–1762. [CrossRef]

54. Everingham, Y.; Sexton, J.; Skocaj, D.; Inman-Bamber, G. Accurate Prediction of Sugarcane Yield Using a Random Forest Algorithm. *Agron. Sustain. Dev.* **2016**, *36*, 1–27. [CrossRef]

55. Jeong, J.H.; Resop, J.P.; Mueller, N.D.; Fleisher, D.H.; Yun, K.; Butler, E.E.; Timlin, D.J.; Shim, K.-M.; Gerber, J.S.; Reddy, V.R.; et al. Random Forests for Global and Regional Crop Yield Predictions. *PLoS ONE* **2016**, *11*, e0156571. [CrossRef] [PubMed]

56. Cheng, S. One-Hundred Years' Development and Prospect of Rice Breeding in China. *China Rice Sci.* **2021**, *27*, 1–6. [CrossRef]

57. Xu, H.; Yin, H.; Liu, Y.; Wang, B.; Song, H.; Zheng, Z.; Zhang, X.; Jiang, L.; Wang, S. Regional Winter Wheat Yield Prediction and Variable Importance Analysis Based on Multisource Environmental Data. *Agron. J.* **2024**, *14*, 1623. [CrossRef]

58. Wang, Q.; Sun, L.; Yang, X. Identifying Spatial Determinants of Rice Yields in Main Producing Areas of China Using Geospatial Machine Learning. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 76. [CrossRef]

59. S, A.; Debnath, M.K.; R, K. Statistical and Machine Learning Models for Location-Specific Crop Yield Prediction Using Weather Indices. *Int. J. Biometeorol.* **2024**, *68*, 2453–2475. [CrossRef] [PubMed]

60. 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] [PubMed]

61. Li, Y.; Tao, F. Rice Yield Response to Climate Variability Diverges Strongly among Climate Zones across China and Is Sensitive to Trait Variation. *Field Crop Res.* **2023**, *301*, 109034. [CrossRef]

62. Devi, S.; Sharma, G.D. Blast Disease in Rice: A Review. *AUJSAT Biomed. Environ. Sci.* **2010**, *6*, 144–154. [CrossRef]

63. Lee, J.; Mast, J.C.; Dessler, A.E. The Effect of Forced Change and Unforced Variability in Heat Waves, Temperature Extremes, and Associated Population Risk in a CO_2-Warmed World. *Atmos. Chem. Phys.* **2021**, *21*, 11889–11904. [CrossRef]