forecasting

MDPI

Article

Rice Yield Forecasting in Northeast China with a Dual-Factor ARIMA Model Incorporating SPEI1-Sep. and Sown Area

Song Nie [1,2] and Zhi-Qiang Jiang [1,2,*]

[1] School of Business, East China University of Science and Technology, Shanghai 200237, China; snie@mail.ecust.edu.cn
[2] Research Center for Econophysics, East China University of Science and Technology, Shanghai 200237, China
* Correspondence: zqjiang@ecust.edu.cn

Highlights

What are the main findings?

- Rural financial development plays a significant role in expanding the land scale for grain farmers by alleviating credit constraints, enhancing mechanization, and supporting land consolidation.
- The impact of agricultural credit on land expansion is particularly strong in maize-dominant regions and small-scale farming households, with mechanization serving as a key intermediary.

What is the implication of the main finding?

- The study highlights the importance of enhancing rural financial infrastructure, especially for smallholder farmers in less-developed areas, to ensure affordable access to credit and mechanization.
- It emphasizes the need for targeted financial policies and interventions in specific regions, particularly in areas with maize farming, to support large-scale farming operations and improve food security.

Abstract

Amid escalating global climate change and geopolitical tensions threatening food supply chains, the three provinces of Northeast China, which serve as a major grain production base, play a crucial role in ensuring national food security. However, the region is experiencing more frequent extreme climatic events and increasing limitations on arable land. This necessitates an evaluation of the combined effects of climate conditions and sown area on rice (*Oryza sativa* L.) yields. Utilizing provincial panel data from 1990 to 2022, this study conducts baseline panel regression analyses at both the national and Northeast China levels. The results consistently identify the value of the standardized precipitation evapotranspiration index (SPEI) on September as a key climatic factor exerting a significant negative effect on rice total yield, whereas the rice sown area is a robust positive determinant. Based on these findings, we develop a dual-factor analytical framework that incorporates both climatic conditions and rice sown area, utilizing SPEI1-Sep. to identify critical growth stages of rice, with the aim of providing a more comprehensive understanding of their combined effects on yield. To further support predictive accuracy, the comparative performance assessments of the Extreme Gradient Boosting (XGBoost), random forest (RF), and Autoregressive Integrated Moving Average (ARIMA) models are conducted. The results show that the ARIMA model outperforms others in forecasting. Forecasts for 2023–2027 indicate slow yield growth in Jilin Province, with a 1.5% annual increase. Heilongjiang shows minor fluctuations, stabilizing between 24.97 and 25.56 million tons. Liaoning's yield remains stable, projected between 5.13 and 5.20 million tons. These

check for updates

Academic Editor: Ignacio A. Ciampitti

Received: 10 September 2025
Revised: 9 November 2025
Accepted: 12 November 2025
Published: 16 November 2025

Citation: Nie, S.; Jiang, Z.-Q. Rice Yield Forecasting in Northeast China with a Dual-Factor ARIMA Model Incorporating SPEI1-Sep. and Sown Area. *Forecasting* **2025**, *7*, 67. https://doi.org/10.3390/forecast7040067

Copyright: © 2025 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/).

trends suggest limited overall yield expansion, highlighting the need for region-specific policies and resource management to ensure China's grain security. This study clarifies the interplay between climate and sown area, demonstrates the relative forecasting advantage of ARIMA in this setting, and provides evidence to support managing yield variability and optimizing agricultural policy in Northeast China, with implications for long-term national food security.

Keywords: food security; SPEI; ARIMA; *Oryza sativa* L.; yield forecasting

1. Introduction

Food security is of paramount national importance, serving as the foundation for economic development, social stability, and national security. As a developing country with a population of 1.4 billion, China has consistently prioritized food security in its governance. However, the current international political landscape is complex and volatile, and the global food security situation is becoming increasingly severe: the ongoing Russia–Ukraine conflict has sharply reduced grain exports from major "world granaries", and many countries have successively implemented export restriction policies, putting global food supply chains at risk of disruption [1–3]. Concurrently, climate change has intensified the frequency of extreme weather events, compounded by fluctuations in global fertilizer and energy prices due to geopolitical conflicts, presenting unprecedented multi-faceted challenges to agricultural production [4–6].

In this context, ensuring stable domestic production and supply of staple grains has been essential for China's national security [7]. Among China's three staple grains, rice plays a key role in ensuring food security. The three provinces of Northeast China (Heilongjiang, Jilin, and Liaoning) account for nearly 20% of the country's total rice yield, making them crucial for maintaining national food supply [8]. However, under the backdrop of global climate change, the threats of extreme climatic events such as droughts and floods are intensifying. Coupled with limited arable land reserves and significant policy-driven adjustments in cropping structures in Northeast China, the regional rice production system faces dual pressures from climate and resource constraints [9,10]. In addition, the sown area, a key variable reflecting policy interventions and market responses, directly influences total yield through its fluctuations across years [11]. Given the combined effects of policies such as the arable land red line, grain-to-green conversion, and agricultural subsidies, changes in sown area have become an essential factor in yield forecasting [12,13].

In the study of the relationship between climate and crop production, the academic community has made substantial progress [14–16]. In particular, in the field of drought monitoring, researchers have proposed a variety of indices, including the Palmer Drought Severity Index (PDSI) based on soil moisture balance [17], the Standardized Precipitation Index (SPI) [18], remote-sensing-based vegetation health index [19], vegetation temperature condition index [20], and other drought indices [21,22]. PDSI and SPI are the most widely used indicators for global drought studies [17,18], however, they have certain limitations. PDSI is suitable only for describing agricultural drought through soil moisture deficits and cannot identify meteorological, hydrological, or socio-economic drought [23]. Moreover, the high dependence of PDSI on data calibration limits the spatial comparability of drought assessments [24]. Although SPI can be used to monitor and evaluate multiple drought types across various spatial scales on a monthly timescale, it considers only precipitation and neglects the evaporative impacts induced by temperature and other meteorological factors [25]. Moreover, when it comes to rice production, most existing studies still focus

on a single climate indicator [26,27], such as precipitation or temperature, without systematically assessing the combined effects of water and heat. To overcome these limitations, Vicente-Serrano et al. [28] integrated the strengths of PDSI and SPI to develop the standardized precipitation evapotranspiration index (SPEI) for drought monitoring and assessment. This index not only addresses the traditional indicators' neglect of temperature effects but also offers the advantage of cross-regional comparability, enabling the characterization of multiple drought types and their spatiotemporal evolution [23]. SPEI has gradually become an essential tool in agricultural climate research due to its comprehensive advantages and has been widely applied in evaluating drought variability, assessing drought impacts, and developing monitoring systems [29–31]. However, research targeting the unique climatic conditions of the three provinces in Northeast China has been relatively scarce, especially concerning the response of rice to climate stress during critical growth stages such as the grain-filling period. This gap in research underscores the need for more focused studies on how climate factors affect rice production in this region, especially during its most vulnerable growth phases.

In addition to the existing gap in climate research, challenges remain in accurately forecasting rice yields in Northeast China, particularly due to the difficulty of integrating climate factors into forecasting models. While machine learning methods, including random forest (RF) [32], extreme gradient boosting (XGBoost) [33], autoregressive integrated moving average (ARIMA) model [8], and long short-term memory networks (LSTM) [34] have advanced yield forecasting models, existing studies have largely focused on the national scale or rice-growing regions in southern China, with insufficient analysis of the regional heterogeneity in the three provinces of Northeast China, a critical food production area. This emphasis on national-scale studies and the lack of regional heterogeneity analysis contribute to the neglect of critical factors such as sown area, a key predictor of yield. Sown area, influenced by natural conditions and policies, is often overlooked in current models. As a result, the integration of climatic factors and cultivation dynamics remains incomplete, offering opportunities to enhance forecast accuracy and stability, which are crucial for more reliable yield predictions [35].

Building on these gaps, this study aims to address three key research needs. First, while many climate-related studies focus on a single factor, few systematically assess the combined water–heat effects under the unique conditions of Northeast China. This gap highlights the need for a more comprehensive approach that integrates both water and heat in the region's agricultural systems. Second, despite its significance, the role of sown area as a core predictor in yield forecasting has often been overlooked, even though it is influenced by natural conditions, policies, and market forces. Finally, existing forecasting models often overlook the regional heterogeneity of Northeast China by focusing solely on generalized climate indicators and sown area, without adequately considering the diverse local conditions and their interactions. To fill these gaps, this research presents three key highlights: (1) it quantifies the physiological response of rice to the combined water and heat effects of SPEI during critical growth stages, (2) it clarifies the impact of sown area on regional yield variations, and (3) it develops an integrated forecasting framework that combines climate and sown area data, using multiple models to optimize yield predictions for Northeast China. By integrating these factors, this study develops a relatively accurate forecasting method that evaluates model performance, ultimately providing a scientific basis for formulating differentiated agricultural policies in the rice-growing regions of Northeast China.

2. Data and Methods

2.1. Data Source

The data on rice yield, sown area, and other agricultural inputs were primarily obtained from the National Bureau of Statistics of China and data from provincial statistical yearbooks and publicly accessible databases of the Ministry of Agriculture and Rural Affairs of the People's Republic of China, covering annual observations from 1990 to 2022. Using panel data from 31 provinces and municipalities (excluding Hong Kong, Macao, and Taiwan), specifically, the datasets contain information on rice output, sown area, and other agricultural production factors. Meteorological data, including monthly precipitation and potential evapotranspiration, were sourced from the 1 km resolution monthly datasets for precipitation and potential evapotranspiration in China, released by the National Tibetan Plateau Data Center. These datasets were collected from weather stations across each province and averaged to provide a comprehensive view of regional climate conditions. The datasets have undergone rigorous quality control and were previously validated in the related studies [36,37]. To address minor data gaps within this period, missing values were supplemented using mean imputation, ensuring the completeness and continuity of the data series.

2.2. Research Method

2.2.1. Standardized Precipitation Evapotranspiration Index

Drought is a significant climatic factor influencing agricultural production in the context of climate change, particularly impacting the growth, development, and yield formation of water-sensitive crops like rice. To systematically assess the spatiotemporal evolution of drought, various indices have been developed, including the Standardized Precipitation Index (SPI), the Palmer Drought Severity Index (PDSI), and the Standardized Precipitation Evapotranspiration Index (SPEI).

Among these, SPEI offers notable advantages due to its unique physical mechanism and comprehensive performance. By combining precipitation with potential evapotranspiration (PET), SPEI overcomes the limitation of SPI, which neglects temperature effects, and better captures the water supply–demand balance in a warming climate. Furthermore, its multi-temporal scale analysis capability (SPEI1–SPEI12) enables diverse applications, ranging from short-term drought monitoring during crop growth seasons to long-term hydrological drought assessments at the watershed scale. Additionally, SPEI can be reliably calculated using only two fundamental meteorological variables, precipitation and temperature, making it highly applicable in regions with limited observational data. Crucially, by standardizing probability distributions, SPEI reduces the influence of regional climatic conditions on drought assessment, providing an objective and quantitative basis for cross-regional drought characterization.

SPEI is a drought indicator developed as an enhancement to SPI, offering a more comprehensive assessment of drought severity by incorporating temperature effects. Unlike SPI, which relies solely on precipitation data, SPEI utilizes a climatic water balance approach, calculating the difference between precipitation (P) and reference evapotranspiration (ET_0) as $D = P - ET_0$. This method enables a more accurate characterization of drought in regions with high temperatures and elevated evaporative demand, and it demonstrates enhanced applicability under a warming climate [28]. The computation of SPEI involves several steps. First, potential evapotranspiration (PET) must be estimated. The original SPEI formula recommended using the Thornthwaite equation for PET calculation [38], as it requires only monthly average temperature and latitude data, making it operationally

simple. However, to improve accuracy, the FAO-recommended Penman–Monteith equation has been widely adopted [39–41], which integrates multiple climatic variables:

$$PET = \frac{0.408\Delta(R_n - G) + \lambda \frac{900}{T+273} u_2 (e_s - e_a)}{\Delta + \lambda(1 + 0.34u_2)} \quad (1)$$

where Δ is the slope of the saturation vapor pressure–temperature relationship (kPa °C^{-1}), R_n is the net radiation at the surface (MJ m^{-2} d^{-1}), and G is the soil heat flux (MJ m^{-2}). λ is the psychrometric constant (kPa °C^{-1}), T is the daily mean air temperature at 2 m height (°C), u_2 is the wind speed at 2 m height (m s^{-1}), e_s is the saturation vapor pressure of air (kPa), and e_a is the actual vapor pressure of air (kPa).

Next, the climatic water balance ($D = P - PET$) is calculated, and the data are normalized using a log-logistic probability distribution, which can be rewritten as:

$$f(z) = \frac{(z-\gamma)\beta}{\alpha^2}\left[1 + \left(\frac{z-\gamma}{\alpha}\right)\right]^{-2} \quad (2)$$

where α, β and γ represent the scale, shape, and origin parameters, respectively. Thus, the cumulative probability distribution function is in the following:

$$F(z) = \left[1 + \left(\frac{\alpha}{z-\gamma}\right)^{\beta}\right]^{-1} \quad (3)$$

Vicente-Serrano et al. [28] defined the formula for the SPEI, which is as follows:

$$SPEI = W - \frac{C_0 + (C_1 + C_2 W)W}{[\delta_1 + (\delta_2 + \delta_3 W)W]W + 1} \quad (4)$$

where $W = \sqrt{-2\ln(P)}$ for $P \leq 0.5$ and $W = \sqrt{-2\ln(1-P)}$ for $P > 0.5$. The constants are $C_0 = 2.515$, $C_1 = 0.802$, $C_2 = 0.010$, $\delta_1 = 1.432$, $\delta_2 = 0.189$, $\delta_3 = 0.001$.

The SPEI enables multi-temporal analysis by calculating values across various timescales, including monthly (SPEI1), quarterly (SPEI3), semi-annual (SPEI6), and annual (SPEI12) periods. This capability supports a wide range of applications, from short-term drought monitoring to the assessment of long-term climatic trends. Additionally, the SPEI provides a clear quantification of hydroclimatic conditions through standardized values: negative values indicate drought, positive values represent wet conditions, and the absolute magnitude reflects the intensity of the event. According to the classification system of the China Meteorological Administration, SPEI values are categorized into seven levels [42], as shown in Figure 1. While specific crops, such as rice, may have distinct water stress thresholds, the generalized SPEI categories are widely used due to their established correlation with rice water demand during critical growth stages [43–45].

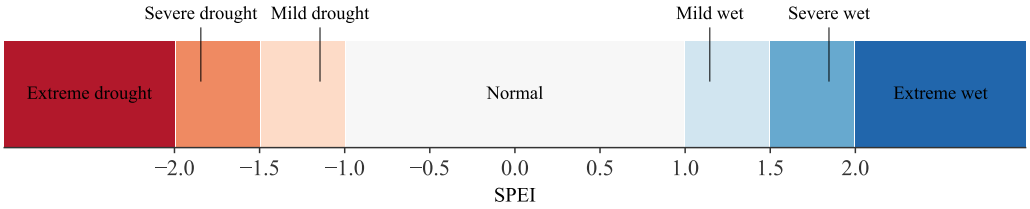


Figure 1. Classification of dry based on SPEI values.

In the subsequent calculations, the SPEI values are derived using monthly precipitation and potential evapotranspiration, as calculated by Equations (1)–(4). In China's rice-growing regions, the rice growth cycle typically lasts 3 to 6 months, depending on

cultivar type, planting location, and cultivation practices. Rice exhibits varying sensitivities to water stress at different growth stages: short-term water deficits (within one month) can cause irreversible damage during critical stages such as tillering, booting-heading, and grain filling, ultimately reducing final yield [28]. This characteristic underscores the unique advantage of the monthly SPEI1 index in analyzing rice's drought responses. It not only enables precise identification of drought conditions at specific growth stages but also facilitates a quantitative assessment of how drought during different stages affects final yield [18]. By leveraging this capability, integrating the SPEI1 index into agricultural economic models can improve the accuracy of yield forecasts, providing more robust scientific support for drought risk assessment and disaster early warning in rice production [16].

2.2.2. Baseline Panel Regression Model

Multiple regression analysis effectively isolates the independent effects of various variables by establishing a linear relationship between multiple independent variables and a dependent variable. The key advantages of this method are reflected in three main aspects: first, by calculating partial regression coefficients, it can accurately assess the net effect of each independent variable on the dependent variable while controlling for other variables; second, the model can effectively capture synergistic effects among variables, addressing potential spurious correlations that may arise from simple correlation analysis; third, the method offers strong scalability, allowing for the flexible incorporation of adjustments such as fixed effects and clustered standard errors. In panel data analysis, multiple regression enhances the accuracy of parameter estimation by simultaneously controlling for the dual fixed effects of individual characteristics and time trends, providing a reliable statistical foundation for causal inference. To systematically assess the impact of drought at different growth stages on rice yield while controlling for confounding effects from agricultural inputs and natural conditions, this study constructs a panel regression model that integrates monthly climate indicators with multidimensional control variables. In the baseline regression analysis, we employ a multivariate fixed-effects model, incorporating the SPEI1 indices for each month from April to September into the regression framework. To address endogeneity issues arising from omitted variables, the baseline regression model controls for both year fixed effects (α_t) and province fixed effects (λ_t).

The α_t capture the effects of unobserved factors that vary over time but are constant across provinces, such as national policies, technological advances, or macroeconomic shocks. These effects are essential in isolating the influence of time-varying factors on rice yield, independent of other factors. The λ_t control for time-invariant unobserved heterogeneity that may differ across provinces, such as regional agricultural practices, geographical conditions, or long-term infrastructure differences. These fixed effects ensure that we do not attribute the variation in rice yield to province-specific unmeasured factors, allowing for more accurate estimations of the impact of the control variables and climatic factors. In addition to the OLS model, which is used to enhance the robustness of estimation results by accounting for these fixed effects, the constructed model is as follows:

$$Y_{m,it} = \beta_0 + \beta_1 SPEI1_{m,it} + \sum \phi_k C_{k,it} + \alpha_i + \lambda_t + \dot{o}_{it} \quad (5)$$

where $Y_{m,it}$ denotes the rice yield (10,000 tons) of province i for month m in year t, $SPEI1_{m,it}$ is the key explanatory variable, representing the $SPEI1$ value for month m in province i and year t, $C_{k,it}$ refers to the k-th control variable in province i and year t. Specifically, $C_{1,it}$ represents the total power of agricultural machinery (10,000 kW), $C_{2,it}$ denotes rural electricity consumption (100 million kWh), $C_{3,it}$ stands for pesticide usage (10,000 tons), $C_{4,it}$ is the total reservoir storage capacity (100 million cubic meters), $C_{5,it}$ refers to the rice sown area (1000 hectares), and $C_{6,it}$ represents the rice disaster-affected areas (1000 hectares).

β_1 is the coefficient for $SPEI1_{m,it}$, representing the marginal effect of the $SPEI1$ index on rice yield, capturing how changes in climatic conditions during month m influence the rice yield in province i in year t. β_0 and $\dot{o}_{it}$ represent the intercept term and the error term, respectively.

Based on the above model specification, Table 1 systematically describes the variable system constructed in this study, including the explained variable (rice yield), the core explanatory variable (monthly-scale SPEI index), and two major categories of control variables covering agricultural production inputs and natural conditions.

Table 1. Variable names and definitions.

Variables	Meanings
Rice_yield	Actual rice yield by province (unit: 10,000 tons)
SPEI1	SPEI at monthly scale (negative values indicate drought, positive values indicate wetness)
Rice_sown_area	Actual sown area for rice planting by province (unit: 1000 hectares)
Reservoir_cap	Total water storage capacity of reservoirs in each province (unit: 100 million cubic meters)
Machinery_power	Total power of agricultural machinery and equipment in each province (unit: 10,000 kilowatts)
Pesticide_use	Total usage of chemical pesticides including insecticides, fungicides, and herbicides in each province (unit: tons)
Rural_elec	Total electricity consumption for production and daily life in rural areas of each province (unit: 100 million kWh)
Disaster_area	Cumulative area of crop yield reduction caused by natural disasters (droughts and floods) in each province (unit: 1000 hectares)

Notes: The data for variable SPEI1 can be referenced from [39] and the National Tibetan Plateau Data Center (https://data.tpdc.ac.cn/home (accessed on 1 November 2025)). Data for other variables are obtained from the National Bureau of Statistics of China (http://www.stats.gov.cn/sj/ (accessed on 1 November 2025)), the Ministry of Agriculture and Rural Affairs of the People's Republic of China (http://www.moa.gov.cn/govpublic/ (accessed on 1 November 2025)).

2.2.3. Multiple Forecast Models

To comprehensively evaluate the applicability and stability of different methods in rice yield forecast, this study selected three representative forecasting models: XGBoost [33], RF [32], and the ARIMA [46]. These three models, respectively, represent gradient boosting methods and bagging methods in ensemble learning, as well as traditional time series modeling approaches. Through comparative analysis of multiple models, the study aims to select the optimal forecasting tool to provide reliable methodological support for subsequent rice yield forecasting.

XGBoost [33] is an ensemble learning algorithm based on the gradient boosting framework, which constructs a strong forecasting model by iteratively optimizing differentiable loss functions. Its core advantage lies in efficiently handling nonlinear relationships and feature interactions. Given a training dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$, where x_i is the feature vector and y_i is the target label, the objective function of the model at the t-th iteration is:

$$L^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)}\right) + \Omega(f_t) \tag{6}$$

where y_i represents the actual value for the ith observation, and $\hat{y}_i^{(t-1)}$ denotes the predicted value of the ith observation from the previous iteration, $t-1$. The term $l(y_i, \hat{y}_i^{(t-1)})$ represents the loss function, which quantifies the error between the actual and predicted values. The function f_t refers to the model's function at the tth iteration. $\Omega(f_t)$, defined as

$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda\|\omega\|^2$, is the regularization term, which helps prevent overfitting by penalizing the complexity of the model. The hyperparameters γ and λ control the complexity of the model, with γ regulating the tree complexity and λ affecting the L2 regularization of the model.

To efficiently optimize the objective function, XGBoost employs a second-order Taylor expansion to approximate the loss function:

$$L^{(t)} \approx \sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)\right] + \Omega(f_t) \tag{7}$$

where g_i and h_i represent the first-order and second-order gradient statistics of the loss function, respectively. The quality of the split points is evaluated through Taylor expansion and structural scores.Through this approximation, XGBoost can quickly evaluate the quality of splitting points, achieving efficient feature selection and model construction.

RF [32] constructs multiple decision trees through Bootstrap sampling and integrates their forecasting results. Its dual randomness effectively reduces the risk of overfitting. For regression tasks, the final forecast value is the mean of all decision tree outputs:

$$\hat{y}(x) = \frac{1}{B}\sum_{b=1}^{B} f_b(x) \tag{8}$$

where $f_b(x)$ is the forecasting function of the b-th tree, and B is the number of trees.

The core principle of RF algorithm lies in its dual randomness. First, Bootstrap sampling, which involves drawing multiple subsets with replacement from the original dataset to train different decision trees. Second, random feature selection, where at each node split, only a randomly selected subset of features is used to find the optimal split point. For regression tasks, the final forecast is obtained by averaging the outputs of all decision trees, whereas for classification tasks, a majority voting mechanism is employed. The RF algorithm inherently resists overfitting, supports parallel training, and provides tools such as out-of-bag error and variable importance for model evaluation and interpretation. This algorithm demonstrates robustness in high-dimensional data, noisy data, and scenarios with missing values.

The ARIMA model [46] transforms non-stationary time series into stationary series through differencing and integrates autoregressive (AR) and moving average (MA) components for modeling. In this paper, for any independent variable $z_t \in f_t, \mathrm{k}_t$, the model is established by a linear time trend with ARIMA(p, d, q) as follows:

$$\varphi(L)(1-L)^d z_t = \beta_0 + \beta_1 t + \theta(L)\varepsilon_t, \varepsilon_t \sim N\left(0, \sigma^2\right) \tag{9}$$

where $\phi(L) = 1 - \sum_{i=1}^{p}\phi_i L^i$ is the autoregressive polynomial, $\theta(L) = 1 + \sum_{j=1}^{q}\theta_j L^j$ is the moving average polynomial, L is the lag operator, d is the differencing order, and ε_t is white noise following a normal distribution. The model employs the augmented Dickey–Fuller test for stationarity assessment, utilizes the autocorrelation function and partial autocorrelation function to identify the orders p and q, and estimates parameters via maximum likelihood estimation.

In practical applications, the performance of the ARIMA model highly depends on parameter selection. This study employs a grid search method to identify the optimal parameter combination within the ranges $p \in [0, 3], d \in [0, 2], q \in [0, 3]$. To ensure the optimality of the model selection, the baseline approach was complemented with a systematic grid search. The search ranges were defined as follows: the autoregressive order (p) and moving average order (q) were explored from 0 to 3, while the differencing order (d) was

set to 1, as confirmed by the Augmented Dickey–Fuller test, which indicated that first-order differencing was sufficient to achieve stationarity. Model evaluation was subsequently conducted using information criteria such as the AIC (Akaike information criterion) and the BIC (Bayesian information criterion).

The model is trained on data from 1990 to 2015 and validated using actual observations from the 2016–2022 period. It is subsequently refitted with the complete dataset spanning 1990 to 2022 to generate five-step-ahead forecasts. The predicted values of the required variables are ultimately derived through linear regression:

$$Y_t = \alpha + \beta_1 f_t + \beta_2 k_t + \eta_t \tag{10}$$

To quantitatively assess the forecasting performance of each model, this study adopts the Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) as evaluation metrics, providing a comprehensive assessment of forecasting accuracy from both relative and absolute error dimensions. Specifically, MAPE measures the average relative deviation between and actual observed values, while RMSE reflects the absolute dispersion of forecasting errors and their sensitivity to outliers.

The formula for $MAPE$ is expressed as:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{11}$$

The formula for $RMSE$ is shown below:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{12}$$

where y is the actual observed value, $\hat{y}$ is the model forecasting value, and n is the sample size.

3. Empirical Results and Mechanism Analysis

The grain production system is a complex system resulting from natural conditions and agricultural technology. This study uses provincial panel data from 1990 to 2022 to identify factors influencing China's grain yield. The dependent variable is rice yield with explanatory variables including machinery power, pesticide use, and rural electricity consumption. The natural conditions include reservoir capacity, rice sown area, agricultural disaster-affected area, and SPEI, shown as Table 1. To analyze drought effects during different growth periods, a panel regression model integrates monthly SPEI indicators with control variables, and empirical tests are conducted at both the national and regional levels. To analyze drought effects during different growth periods, a panel regression model integrates monthly SPEI indicators with control variables, and empirical tests are conducted at both the national and regional levels of the three provinces in Northeast China using Stata 17.0, as specified in Equation (5). Different variations of the SPEI index are modeled: Model 1 uses SPEI1-Apr., Model 2 uses SPEI1-May, Model 3 uses SPEI1-Jun., Model 4 uses SPEI1-Jul., Model 5 uses SPEI1-Aug., and Model 6 uses SPEI1-Sep. Each model captures the impact of different temporal scales of SPEI on rice yield.

3.1. Baseline Panel Regression Results at the National Level

At the national level, this study empirically analyzes the impact of the SPEI on rice yield using panel data from 31 provinces and municipalities (excluding Hong Kong, Macao, and Taiwan). The model is specified in Equation (5). The rice yield is modeled with different variations of the SPEI index for the national level in China, where the core explanatory

variable changes across models: Model 1 uses SPEI1-Apr., Model 2 uses SPEI1-May, Model 3 uses SPEI1-Jun., Model 4 uses SPEI1-Jul., Model 5 uses SPEI1-Aug., and Model 6 uses SPEI1-Sep.. Each model thus captures the influence of different temporal scales of the SPEI on rice yield. The results show that the maximum variance inflation factor (VIF) is 4.46, with a mean of 2.05, indicating no multicollinearity issues among the explanatory variables. Table 2 presents the baseline regression results at the national level.

Table 2. Regression results at the national level.

Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
SPEI1-Apr.	−0.354					
	(−0.19)					
SPEI1-May		6.990 ***				
		(3.29)				
SPEI1-Jun.			0.949			
			(0.54)			
SPEI1-Jul.				1.257		
				(0.57)		
SPEI1-Aug.					−2.303	
					(−1.06)	
SPEI1-Sep.						−3.204 **
						(−2.17)
Rice_sown_area	0.681 ***	0.680 ***	0.681 ***	0.681 ***	0.681 ***	0.681 ***
	(18.86)	(19.33)	(18.86)	(18.85)	(18.84)	(18.94)
Machinery_power	0.007	0.007	0.007	0.007	0.007	0.007
	(0.93)	(0.95)	(0.92)	(0.93)	(0.92)	(0.95)
Disaster_area	−0.035 ***	−0.035 ***	−0.035 ***	−0.036 ***	−0.035 ***	−0.035 ***
	(−6.50)	(−6.32)	(−6.35)	(−6.29)	(−6.45)	(−6.56)
Rural_elec	0.044 **	0.043 **	0.044 **	0.044 **	−0.044 **	−0.044 **
	(2.04)	(2.00)	(2.04)	(2.04)	(2.05)	(2.08)
Reservoir_cap	0.108 *	0.107 *	0.108 *	0.109 *	0.108 *	0.109 *
	(1.74)	(1.74)	(1.73)	(1.75)	(1.73)	(1.75)
Pesticide_use	6.743 *	6.752 *	6.783 *	6.781 *	6.824 *	6.604 *
	(1.78)	(1.77)	(1.78)	(1.78)	(1.78)	(1.75)
Constant	−141.382 ***	−143.717 ***	−141.538 ***	−141.710 ***	−142.595 ***	−141.204 ***
	(−3.17)	(−3.31)	(−3.19)	(−3.19)	(−3.20)	(−3.19)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1023	1023	1023	1023	1023	1023

Notes: The statistical significance of the regression coefficients was tested using *t*-tests. ***, **, and * represent 1%, 5%, and 10% significance level, respectively, and *t*-statistics in parentheses.

Statistical analysis was performed using Stata 17.0 (StataCorp LLC, College Station, TX, USA), and a panel regression model with Fixed Effects was employed. The statistical significance of the coefficients was assessed using t-statistics, with p-values used to determine the significance level. The model controlled for both time and provincial fixed effects to ensure robust estimation of the relationships between the variables. From the regression results, it can be observed that rice sown area and rural electricity consumption have a significant positive impact on rice yield, while disaster-affected area exhibits significant negative effects.The impact of Machinery power on rice yield is statistically insignificant. Additionally, pesticide usage and reservoir capacity show weaker impacts, being significant only at the 10% level.

It is worth emphasizing that this study does not distinguish among early-, mid-, and late-season rice, because the April-to-September window broadly spans the dominant growth cycle of rice across China and is therefore climatically decisive for grain formation.

In official statistics, rice is conventionally classified by cropping season into early rice, mid-season rice, single-season rice and late rice, with mid-season and single-season rice accounting for roughly 75% of national output (National Bureau of Statistics of China & Ministry of Agriculture and Rural Affairs of the People's Republic of China). As the principal ripening phase of these cultivars occurs mainly in September–October, focusing on April–September captures all critical phenophases from transplanting through grain filling within a single analytical frame. A comprehensive assessment of climatic conditions during this interval thus enables the identification of the key weather drivers of yield variability, supports more accurate rice-yield projections, and provides an evidence base for formulating targeted policies that safeguard national food security.

Notably, SPEI1-May (May SPEI) demonstrates a significant positive relationship with rice yield, whereas SPEI1-Sep. (September SPEI) shows a significant negative impact. This result indicates that higher water availability in May benefits rice growth during the tillering stage, while excessive water in September may inhibit grain filling during the ripening stage, which aligns with the water demand characteristics of rice at different growth stages. At the national level, May coincides with a critical growth period for rice, particularly the transplanting and recovery stage in southern rice regions, where moderate water conditions are crucial for seedling survival and tiller formation. The significant positive association of SPEI1-May suggests that higher water availability in May effectively promotes root development and population establishment. Firstly, sufficient precipitation alleviates seasonal drought stress on seedlings in spring. Secondly, the lower evapotranspiration demand reflected by moderate positive SPEI reduces water loss, maintaining optimal water depth in paddy fields and creating favorable conditions for tiller formation. Furthermore, the negative significant relationship between rice yield in September and SPEI1-Sep. is related to changes in water demand during crop growth stages. In most rice regions, September marks the grain-filling and maturation phase, where moderate drought actually benefits dry matter accumulation and grain filling.

Excessively high SPEI conditions may trigger adverse effects: (1) overcast and rainy weather reduces photosynthetically active radiation, slowing grain-filling rates, (2) high-humidity environments increase pest and disease risks, and (3) excessive rainfall may lead to lodging, particularly for taller varieties like hybrid rice. Thus, the negative association in September reflects the sensitivity of reproductive growth stages to water surplus.

3.2. Baseline Panel Regression Results for the Three Provinces of Northeast China

The models are calculated by using Stata 17.0, as specified in Equation (5). The rice yield is modeled with different variations of the SPEI index for the three Provinces of Northeast China, where the core explanatory variable changes across models: Model 7 uses SPEI1-Apr., Model 8 uses SPEI1-May, Model 9 uses SPEI1-Jun., Model 10 uses SPEI1-Jul., Model 11 uses SPEI1-Aug., and Model 12 uses SPEI1-Sep. Analysis of data from the three Provinces of Northeast China shows that the maximum VIF does not exceed 10, with a mean of 3.45, further confirming the absence of multicollinearity in the model. Table 3 reports the baseline regression results for the three provinces of Northeast China.

In the three provinces of Northeast China, June and September correspond to critical stages in rice growth and development, with environmental conditions during these months exerting a decisive influence on final yield. Rice yield in this region shows a significant negative correlation with SPEI1-Jun. (June SPEI), a relationship closely tied to the region's unique climatic conditions and the characteristics of the rice-growing season. Northeast China follows a single-cropping pattern, with June marking the rice tillering stage. Unlike southern rice-growing areas, spring warming in this region is relatively slow, while precipitation in June is generally abundant. Higher SPEI1-Jun. levels may have

three primary adverse effects. First, low temperatures and reduced sunlight can inhibit tiller formation, delaying the overall growth process. Second, overly wet soils impair root activity, reducing nutrient uptake efficiency. Finally, high humidity significantly increases the risk of diseases, such as rice blast. In contrast, moderate water deficits can promote tillering by enhancing soil temperature, increasing light availability, and reducing disease incidence. This contrasts with the positive effect of May SPEI observed at the national level.

Table 3. Regression results at the level of the three provinces of Northeast China.

Variables	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
SPEI1-Apr.	6.776					
	(1.03)					
SPEI1-May		6.383				
		(0.38)				
SPEI1-Jun.			−28.239 *			
			(−4.18)			
SPEI1-Jul.				−18.543		
				(−1.51)		
SPEI1-Aug.					−17.883	
					(−2.06)	
SPEI1-Sep.						−35.095 **
						(−7.81)
Rice_sown_area	0.781 ***	0.777 ***	0.757 ***	0.786 ***	0.769 ***	0.766 ***
	(45.08)	(34.88)	(37.49)	(31.25)	(63.33)	(35.50)
Machinery_power	0.008	0.011	0.022	0.007	0.016	0.012
	(0.49)	(0.58)	(1.20)	(0.38)	(1.04)	(0.62)
Disaster_area	−0.013 *	−0.014 *	−0.011 **	−0.013 *	−0.015 **	−0.013 **
	(−3.44)	(−3.28)	(−5.56)	(−4.00)	(−4.48)	(−4.91)
Rural_elec	0.056	0.065	0.050	0.062	0.046	0.021
	(0.59)	(0.66)	(0.51)	(0.61)	(0.49)	(0.25)
Reservior_cap	−0.911	−0.896	−0.692	−0.886	−0.780	−0.724
	(−2.85)	(−2.92)	(−2.16)	(−2.88)	(−2.80)	(−2.33)
Pesticide_use	2.866	2.280	7.003	0.398	1.563	6.729
	(0.18)	(0.14)	(0.41)	(0.02)	(0.11)	(0.40)
Constant	88.726	85.262	59.412	91.218	64.405	82.647
	(1.73)	(1.61)	(1.28)	(2.21)	(1.76)	(1.41)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Province FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1023	1023	1023	1023	1023	1023

Notes: The statistical significance of the regression coefficients was tested using *t*-tests. ***, **, and * represent 1%, 5%, and 10% significance level, respectively, and *t*-statistics in parentheses.

The significantly negative impact of September SPEI1-Sep. further underscores the distinct characteristics of the northeastern rice region. In September, rice enters the grain-filling and maturation stage, but temperatures begin to drop noticeably. Higher SPEI1-Sep. levels limit yield through two main mechanisms: prolonged cloudy and rainy weather reduces accumulated temperature, slows the grain-filling rate, and increases the likelihood of incomplete grain development; sustained autumn rainfall can also lead to poor field drainage and early root senescence. Conversely, moderately dry conditions help raise soil temperature and promote the translocation of photosynthates. Although the negative effect of September SPEI at the national level is similar in direction, the underlying mechanisms differ regionally. In Northeast China, ensuring sufficient accumulated temperature is more critical, while at the national level, the focus is generally on pest and disease control.

This regional variation in the timing and direction of impacts highlights the spatial differentiation patterns of the climate–crop system. In Northeast China, the negative

effect in June emphasizes the antagonistic relationship between water and temperature in cold-region rice yield, while the negative effect in September reinforces the threshold of water management under temperature constraints. These findings suggest that, in Northeast China, measures such as field drying should be implemented during the tillering stage to avoid excessive moisture, while drainage and waterlogging prevention should be prioritized during the maturation stage. Compared with the national level, the SPEI response characteristics in Northeast China are more dependent on thermal regulation, providing a critical foundation for regional adaptive cultivation strategies. These findings indicate that water management strategies in this region may require dynamic adjustments under climate warming.

Additionally, the rice sown area has a significantly positive impact across all models, highlighting its importance as a key structural factor in determining regional total output. Unlike climatic factors, interannual fluctuations in sown area are more influenced by policy adjustments, market returns, and constraints on arable land resources, giving it forecasting value independent of climatic factors. This provides empirical support for constructing a dual-factor forecasting model based on climate and area.

3.3. Mechanisms of Climatic Factors During Key Growth Stages

In the correlation analysis between rice yield and the SPEI at both the national level and in the three provinces of Northeast China, the September SPEI1-Sep. consistently showed significant impacts, although the direction of its effect varied across regions. This commonality suggests that the moisture–heat balance in September plays a critical regulatory role in rice yield formation. September coincides with the late reproductive growth stage of rice, during which most rice-growing regions in China are in the grain-filling to maturity phase. However, due to the colder climate in Northeast China, the rice growth period is relatively delayed. The negative effect of SPEI1-Sep. during this period may result from similar physiological and ecological mechanisms: excessively high SPEI can reduce photosynthetically active radiation, slow the grain-filling rate, and increase the risks of lodging and pre-harvest sprouting. Nevertheless, the negative effect is more pronounced in the three provinces of Northeast China, likely due to the region's rapid autumn cooling. Excessive precipitation in September, combined with low temperatures, further inhibits grain filling, while moderate water deficits may instead help raise soil temperature and promote dry matter translocation. This finding suggests that, despite differences in climatic backgrounds and cropping systems across China's rice-growing regions, the impact of SPEI1-Sep. on rice yield is universal, particularly in regions with greater climate variability, where its forecasting value is more significant.

This study selected the September SPEI1-Sep. in the three provinces of Northeast China as one of the core indicators for rice yield forecasting, based on three key scientific rationales. First, the three provinces belong to a relatively homogeneous single-season rice climate zone, where the spatial and temporal variability of September temperature and precipitation is small. This makes SPEI1-Sep. highly representative for the region and facilitates the construction of a unified forecasting model. Second, rice yield in Northeast China accounts for over 20% of the national total, and its stability is crucial for national food security. September is a critical period for local rice yield formation, and SPEI1-Sep. effectively captures the synergistic effects of water stress and heat limitation during the grain-filling stage. Third, compared to other rice-growing regions in China, the negative effect of September SPEI1-Sep. in Northeast China is more stable, as it is less affected by complex weather phenomena such as typhoon rainfall or autumn rains in western China, resulting in a higher signal-to-noise ratio. The forecasting capability of this indicator is not only reflected in its statistical significance but also in the fact that it can be easily obtained

from meteorological observations or short-term climate forecasts, making it practical for use. Focusing on the forecasting research of September SPEI1-Sep. for rice yield in Northeast China can deepen the understanding of rice climate adaptation mechanisms and provide decision-making support for regional agricultural disaster prevention and mitigation. Based on historical data analysis, if September SPEI1-Sep. in a given year is significantly high, early warnings can be issued for potential grain-filling obstacles, guiding farmers to take preventative measures such as field drainage.

In summary, September SPEI1-Sep. has a significant negative regulatory effect on rice yield, particularly demonstrating strong predictive capability during the grain-filling stage in the three provinces of Northeast China. However, climatic factors alone can explain only part of the yield variation. The sown area, as a key variable reflecting policy adjustments, market responses, and arable land resource constraints, directly and stably contributes to regional total rice yield. In the baseline panel regression analysis, rice sown area showed a significant positive effect at both the national level and in the three provinces, confirming its role as one of the core determinants of total rice yield. Theoretically, a larger sown area generally implies a higher potential harvest. This finding also underscores that sown area and climatic variables are the two key factors influencing rice yield: the former represents an essential production input, while the latter reflects significant natural determinants.

4. Rice Yield Forecast and Validation in the Three Provinces of Northeast China

Northeast China, consisting of Heilongjiang, Jilin, and Liaoning provinces, is a key grain-producing region, contributing approximately 20% of the national rice yield [8,47]. However, limited arable land and climate change pose risks to future production. This section utilizes the SPEI index and rice sown area to forecast rice yields in the three provinces, aiming to support regional food security and policy development.

4.1. Construction of the Dual-Factor Forecasting Index System: Climate and Sown Area

When constructing a rice yield forecasting model for the three provinces of Northeast China, it is essential to comprehensively account for the dual influence of climatic factors and agricultural production characteristics. Based on the regression analysis results from the previous section, this study selects the SPEI1-Sep. and rice sown area as core predictors for the forecasting model.

First, SPEI1-Sep. serves as a comprehensive indicator of the water–heat balance in September, directly influencing the physiological processes of rice during the grain-filling stage. In the three provinces of Northeast China, September is a critical period for rice yield formation. The negative significant effect of SPEI1-Sep. reflects the region's unique agroclimatic characteristics: excessive water availability is often accompanied by low temperatures and reduced sunlight, which delays grain filling, while a mild water deficit can enhance soil temperature and light utilization, thereby promoting grain plumpness. This predictor not only holds clear biological significance but can also be obtained in real time from meteorological data, offering strong operability and timeliness.

Second, the rice sown area is a direct determinant of regional total output, with interannual fluctuations closely linked to policy adjustments, market supply and demand, and changes in land use. In Northeast China, expansions or contractions in the sown area often precede yield changes and significantly contribute to total production. Incorporating sown area into the forecasting model effectively captures the impact of human factors on yield, addressing the limitations of climate-based models alone.

Finally, the combination of SPEI1-Sep. and rice sown area captures the synergy between climate stress and policy-driven production scale: the former reflects the constraints

of environmental stress on yield, while the latter represents the production potential under policy regulation. This dual-factor framework aligns with the complexity of agricultural systems while maintaining practical simplicity, providing robust scientific support for rice yield forecasting in Northeast China.

4.2. Comparison of Different Models and Performance Evaluation

This study, based on the regional characteristics of rice yield in the three provinces of Northeast China, aims to develop a forecasting model for rice yield in the region, providing scientific projections of yield trends from 2023 to 2027. As a key grain production base in China, the three provinces account for approximately 20% of the nation's total rice yield over the past five years. An accurate yield forecast for this region is essential not only for local food security but also for the national food supply–demand balance.

In constructing the model, this study employed a multidimensional comparative analysis to evaluate three representative forecasting models: RF, XGBoost, and ARIMA model. To ensure the comprehensiveness and reliability of the model evaluation, two key metrics, MAPE and RMSE, were systematically applied. MAPE intuitively reflects the relative magnitude of forecasting errors, while RMSE penalizes larger errors. Their combined use provides a comprehensive assessment of model performance.

Regarding data sources and processing, this study collected 33 years of historical data (1990–2022) for the three provinces, including rice yield (in 10,000 tons), rice sown area (in 1000 hectares), and the SPEI1-Sep. index. To ensure adequate model training and reliable testing, the dataset was split into a training set (1990–2016) and a test set (2017–2022) in an 80:20 ratio. Subsequently, MAPE and RMSE were calculated for each model. The model performance evaluation results are presented in Table 4.

Table 4. Performance evaluation of forecasting models in the three provinces of Northeast China (MAPE and RMSE).

	Jilin					**Heilongjiang**					**Liaoning**				
	RF	**XGBoost**	**ARIMA**	**LSTM**	**GRU**	**RF**	**XGBoost**	**ARIMA**	**LSTM**	**GRU**	**RF**	**XGBoost**	**ARIMA**	**LSTM**	**GRU**
MAPE (%)	13.59	8.14	3.86	10.51	9.47	13.30	7.07	2.82	8.52	8.63	16.60	20.03	14.87	16.22	16.17
RMSE (10^4 tons)	122.83	74.10	30.81	86.08	68.89	460.81	248.29	100.22	290.5	287.76	83.44	94.36	64.05	72.27	72.40

The comparative analysis of models revealed that the ARIMA model demonstrated the higher performance in rice yield forecasting across the three provinces of Northeast China, although forecasting accuracy varied by region. In Jilin Province, the ARIMA model achieved the lowest forecasting error with MAPE = 3.86% and RMSE = 30.81 (10^4 tons), significantly outperforming XGBoost and RF. The results for Heilongjiang Province further validated the superiority of the ARIMA model, with both MAPE = 2.82% and RMSE = 100.22 (10^4 tons) being the lowest among the three provinces. This may be attributed to the province's large-scale rice cultivation, higher production stability, and stronger regularity in time-series data. In contrast, the forecasting accuracy of the ARIMA model in Liaoning Province with (MAPE = 14.87% and RMSE = 64.05 (10^4 tons), while still better than that of other models, was lower compared to Jilin and Heilongjiang. This discrepancy may stem from Liaoning's greater climate variability, leading to larger interannual fluctuations in rice yield and increasing the difficulty of time-series forecasting.

In addition, to further enhance the accuracy of rice yield forecasting, we also incorporated LSTM (long short-term memory) [48] and GRU (gated recurrent unit) models [49], which are designed to capture temporal dependencies and nonlinear patterns in time-series data. These deep learning models were applied to compare their performance with the traditional ARIMA model. Both LSTM and GRU models are well-suited for modeling sequential data and can handle long-term dependencies and complex non-linearities that

may be present in the rice yield time-series. The MAPE and RMSE values for these models, as shown in Table 4, illustrate their ability to capture the dynamics of rice yield fluctuations across the three provinces. Comparative analyses indicate that the ARIMA model also outperforms both LSTM and GRU architectures in forecasting accuracy, corroborating its capacity to effectively capture the time-series characteristics of rice yields in Northeast China, particularly in regions where production patterns are more stable and predictable.

The superiority of the ARIMA model in forecasting rice yields in the three provinces of Northeast China can be attributed to several key factors. First, the model effectively handles non-stationary time-series data by eliminating long-term trends and seasonal fluctuations through differencing, thereby more accurately capturing the interannual variation patterns of rice yields. Next, the ARIMA model exhibits strong dynamic responsiveness to climatic factors (SPEI1-Sep.), enabling better simulation of the nonlinear impact mechanisms of drought stress on rice growth. Finally, the model can integrate production factors such as sown area, establishing dynamic correlations between yields and key variables through autoregressive (AR) and moving average (MA) mechanisms, thereby enhancing the robustness of long-term forecasts.

4.3. Rice Yield Forecasting Results for the Three Provinces of Northeast China

This study employs the ARIMA model to construct a two-phase forecasting process, with the specific steps outlined as follows:

Phase one: time-series forecasting of predictive factors. First, based on the ARIMA model, time-series forecasts are made for the rice sown area and climate factors (SPEI1-Sep.) in the three provinces of Northeast China from 2023 to 2027. The changing trends of rice sown areas in these three provinces are shown in Figure 2.

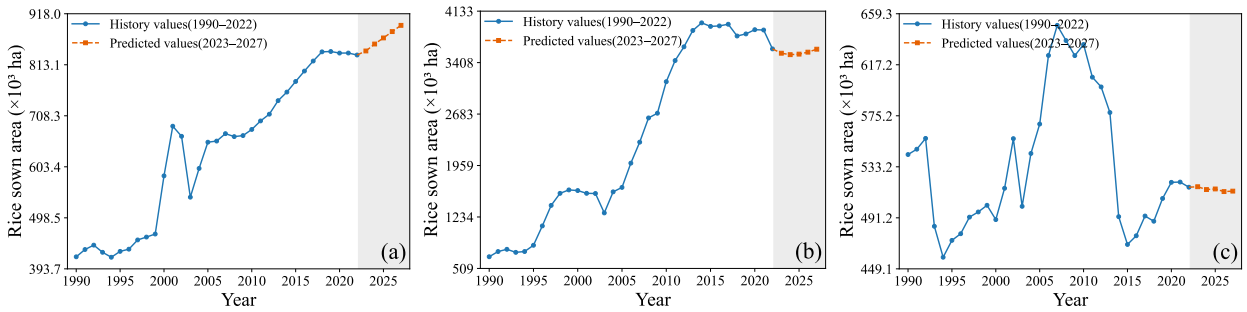


Figure 2. Plots of the trend on rice sown area of Northeast China. (**a**) Jilin. (**b**) Heilongjiang. (**c**) Liaoning.

Based on the statistical and forecasting data of rice sown areas in the three provinces from 1990 to 2027, this study adopts a method combining quantitative analysis and policy evaluation to systematically reveal the dynamic characteristics and driving mechanisms of regional sown area changes. The research finds that changes in rice sown areas in the three provinces exhibit significant phased differences, with 2003 being a critical turning point. That year, the sown areas in Heilongjiang, Jilin, and Liaoning decreased by 17.5%, 19.1%, and 10%, respectively. This synchronous decline was closely linked to the implementation of two national policies: first, the enforcement of the "Regulations on Returning Farmland to Forests" in the northeastern region in 2002, aimed at alleviating ecological pressure by converting some rice fields into forests; second, the planting structure adjustment policy promoted by the Ministry of Agriculture in 2003, which guided farmers to optimize crop layouts. These policies directly impacted the agricultural production structure of the three provinces, reflecting the profound regulatory role of macro policies on regional agricultural development.

Heilongjiang Province's rice sown area peaked in 2014 at 3968.48 × 10^3 hectares, before exhibiting a fluctuating downward trend, reaching 3601.37 × 10^3 hectares in 2022. The ARIMA model-based forecasting results indicate that Heilongjiang's sown area will stabilize at 3520–3540 × 10^3 hectares from 2023 to 2024, with minor fluctuations, and is expected to rebound to 3597.34 × 10^3 hectares by 2027, demonstrating a certain degree of adjustment resilience in its production system. Jilin Province, on the other hand, has maintained steady growth, increasing from 418.4 × 10^3 hectares in 1990 to 833.18 × 10^3 hectares in 2022. Although it experienced a decline to 540.95 × 10^3 hectares in 2003, recovery was rapid thereafter. During the forecasting period of 2023–2027, it is expected to continue a moderate upward trend, reaching 894.16 × 10^3 hectares by 2027, with an average annual growth rate of about 1.5%, indicating strong production resilience. Liaoning Province maintained relatively stable growth with fluctuations before 2010, increasing from 543.3 × 10^3 hectares to 633.93 × 10^3 hectares, with an average annual growth rate of approximately 0.8%. However, from 2011 to 2015, the area dropped from 607.01 × 10^3 hectares to 469.23 × 10^3 hectares, followed by a gradual recovery after 2015. This U-shaped curve was primarily driven by three factors: the farmland restrictions under the "Liao River Basin Comprehensive Management Plan", land occupation due to the development of the Bohai Rim Economic Belt, and declining planting profitability. After 2015, with strengthened farmland protection policies and increased subsidies, the area recovered to 516.39 × 10^3 hectares by 2022 and is expected to stabilize at 500–520 × 10^3 hectares during the forecasting period, marking the entry of its production layout into a mature phase. From a regional perspective, although Heilongjiang Province has a prominent scale advantage, the recent downward trend requires close attention; Jilin Province's sustained growth provides stable support for regional food security; and Liaoning Province has achieved a rebalancing of its production system through policy adjustments, entering a stable development phase.

According to the training and forecasting results of the ARIMA model, the climate factors (SPEI1-Sep.) for Jilin, Heilongjiang, and Liaoning are shown in Figure 3.

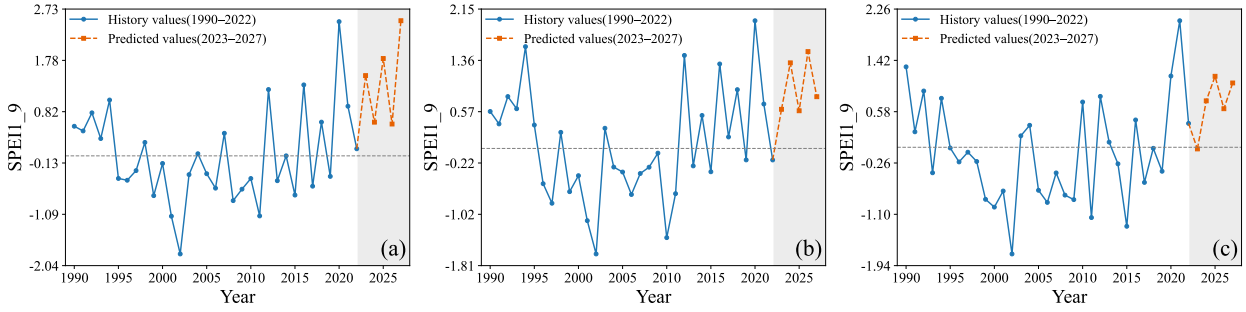


Figure 3. Plots of the trend on SPEI1-Sep. of Northeast China. (**a**) Jilin. (**b**) Heilongjiang. (**c**) Liaoning.

Based on historical data of the SPEI1-Sep. from 1990 to 2022, along with forecast data for 2023–2027, this paper analyzes the regional trends in dryness and wetness variation across the three provinces of Northeast China. The SPEI1-Sep. reflects conditions of dryness and wetness using standardized values: negative values indicate drought, positive values indicate wetness, and the larger the absolute value, the stronger the intensity. The typical growth requirements for rice during its heading and grain-filling stages generally fall within the range of 0 to 1.0, corresponding to mild wet conditions.

In Heilongjiang Province, the SPEI1-Sep. values from 1990 to 2022 remained close to normal levels but exhibited significant fluctuations. The extreme drought years were 2002 (SPEI1-Sep. = −1.63) and 2010 (SPEI1-Sep. = −1.38), while the extreme wet years occurred in 2020 (SPEI1-Sep. = 1.97) and 1994 (SPEI1-Sep. = 1.57). Forecasts for 2023–2027 indicate an increasing trend toward wetter conditions, with all future SPEI1-Sep. values remaining

positive. Notably, 2024 (SPEI1-Sep. = 1.32) and 2026 (SPEI1-Sep. = 1.49) fall within the "very wet" range, while 2027 returns to "normal" levels (SPEI1-Sep. = 0.80), suggesting a potential rise in short-term extreme wet events.

Jilin Province, in contrast, experienced a slightly higher frequency of drought events than Heilongjiang. Severe droughts occurred in 2002 (SPEI1-Sep. = −1.83) and 2001 (SPEI1-Sep. = −1.12), while exceptionally wet years included 2020 (SPEI1-Sep. = 2.50) and 2021 (SPEI1-Sep. = 0.92). The forecast period shows a significant wetting trend, with the annual average value rising to 1.41. The years 2025 (SPEI1-Sep. = 1.81) and 2027 (SPEI1-Sep. = 2.51) reach levels of "very wet" to "extremely wet," which may increase flood risks and suggest the need for caution regarding potential agricultural impacts of high SPEI values.

In Liaoning Province, drought events were frequent from 1999 to 2002 (SPEI1-Sep. ≤ -0.85), with 2021 marking an extreme wet peak (SPEI1-Sep. = 2.07). Forecasts indicate wet conditions in 2025 (SPEI1-Sep. = 1.16) and 2027 (SPEI1-Sep. = 1.05), while other years remain close to normal levels, demonstrating greater overall climate stability compared to Heilongjiang and Jilin.

The SPEI1-Sep. variation characteristics across the three provinces of Northeast China reveal that Jilin has the highest frequency of drought events and the greatest risk of extreme events (e.g., SPEI1-Sep. reaching 2.51 in 2027), Heilongjiang shows the most significant fluctuations between wet and dry conditions, and Liaoning's climate conditions are relatively stable. During the forecast period (2023–2027), all three provinces exhibit a trend toward wetter conditions.

Phase two: rice yield forecast. Using the forecasting values of the rice sown area and climate factor SPEI1-Sep. obtained in Phase One as input variables, the ARIMA model calculates the forecasted rice yields for each province, as shown in Figure 4.

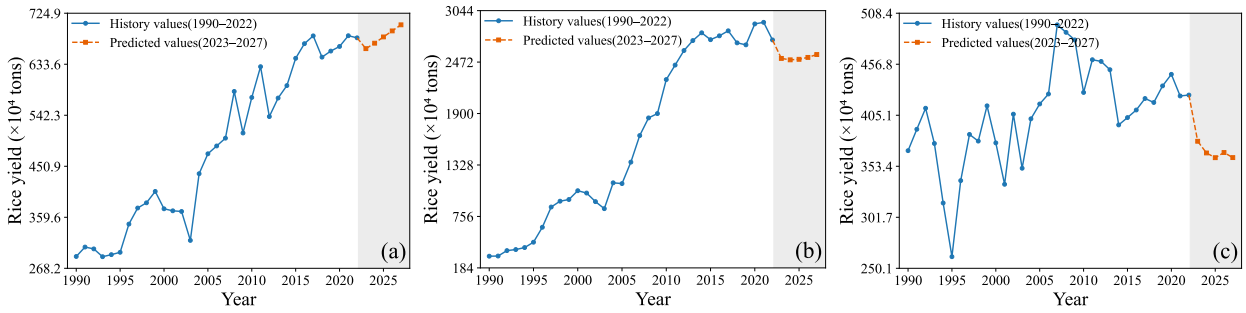


Figure 4. Plots of the trend on rice yield of Northeast China. (**a**) Jilin. (**b**) Heilongjiang. (**c**) Liaoning.

To evaluate the performance of the ARIMA model in forecasting rice yields in Northeast China, this study first validated it using actual yield data from 2023 to 2024. In 2023, the actual rice yields for Jilin, Heilongjiang, and Liaoning were 6.8206 million tons, 24.4000 million tons, and 4.1234 million tons, respectively. The corresponding ARIMA model forecasts were 7.0312 million tons, 25.1272 million tons, and 4.2012 million tons, with relative errors of 3.01%, 2.98%, and 8.15%, respectively. In 2024, the actual yields were 6.7638 million tons, 24.7160 million tons, and 3.9840 million tons, while the forecasted values were 6.7131 million tons, 24.9676 million tons, and 3.6685 million tons, with relative errors of 0.75%, 1.02%, and 7.92%, respectively. Over the two years, all other errors remained below 9%, meeting agricultural forecasting accuracy requirements and confirming the ARIMA model's reliability and applicability for rice yield forecasting in Northeast China.

Based on the confirmed forecasting performance of the model and further analysis of historical data, it is evident that rice yield in the three provinces of Northeast China exhibits distinct regional disparities. Heilongjiang Province, as the primary production base, shows the most significant growth trajectory, with yield increasing from 3.144 million tons in

1990 to 27.18 million tons in 2022, representing an average annual growth rate of 7.8%. After peaking at 27.9722 million tons in 2014, the province experienced a slight decline, a trend consistent with its reduction in sown area. Jilin Province demonstrates steady growth, maintaining an average annual growth rate of around 3.1%, with synchronized expansion in both yield and sown area, reflecting the stability of its agricultural production system. Liaoning Province displays typical policy-responsive characteristics, with yield declining from 6.3393 million tons to 4.6923 million tons between 2010 and 2015 due to the influence of the "Liao River Basin Comprehensive Management Plan," before gradually recovering to 5.1639 million tons by 2022, forming a complete U-shaped adjustment curve. Notably, 2003 was a critical year for all three provinces, with significant yield declines: Heilongjiang dropped by 17.5%, Jilin by 19.1%, and Liaoning by 10%. This phenomenon aligns closely with reductions in the rice sown area and is linked to policy adjustments and planting structure optimization at the time. Among the three provinces, Jilin exhibited the fastest recovery, while Liaoning was the slowest, highlighting differences in the adaptability of regional agricultural production systems to policy changes. This model-supported differential analysis lays the foundation for understanding the driving mechanisms behind yield variations in each province.

Furthermore, combined with SPEI1-Sep. forecasting data, the three provinces of Northeast China are projected to enter a humid climate period from 2023 to 2027. Heilongjiang's extreme wet years (SPEI1-Sep. = 1.32 in 2024, SPEI1-Sep. = 1.49 in 2026) coincide with forecasting yield lows (24.9676 million tons in 2024), suggesting excessive moisture may lead to reduced yields. Despite Jilin's highest SPEI1-Sep. value (reaching 2.51 in 2027), its yield continues to grow. Liaoning's SPEI1-Sep. fluctuations are relatively minor, and its corresponding yield forecasts show high stability, validating the protective role of climatic stability in agricultural production.

Forecasts based on the ARIMA model indicate that Heilongjiang Province is expected to experience a slight decline in production before stabilizing, with projected values for 2023–2027 ranging between 24.9676 and 25.5563 million tons, aligning with the low-level fluctuation characteristic of its cultivated area, suggesting limited potential for yield improvement. In contrast, Jilin Province continues its growth trajectory, with yield projected to reach 7.0419 million tons by 2027, accompanied by an increase in cultivated area to 894.16×10^3 hectares, reflecting an average annual growth rate of 1.5%. This demonstrates its ability to maintain yield resilience under high SPEI wet conditions. Liaoning Province has entered a relatively stable phase of rice yield development, with projected yield for 2023–2027 remaining between 5.13 and 5.20 million tons, while its cultivated area stabilizes within the range of $500–520 \times 10^3$ hectares. This trend indicates that the province's agricultural production system is approaching dynamic equilibrium, consistent with the characteristics of its mature policy regulation phase.

In summary, the forecasts for the three provinces of Northeast China reveal distinct trajectories that carry significant implications for regional and national food security. Heilongjiang's slight production decline and limited cultivated area expansion suggest constrained yield improvement potential, highlighting a need for targeted technological or policy interventions to prevent stagnation. Jilin's sustained growth in both production and cultivated area underscores its role as a stabilizing force in regional rice output, demonstrating resilience under favorable climatic conditions. Liaoning's plateauing production reflects a mature and regulated agricultural system, indicating stability but limited capacity for further yield gains. Collectively, these trends suggest that while northeastern China will continue to contribute substantially to national rice supply, the overall growth potential is uneven, and any adverse climatic events or policy shifts could disproportionately affect total output. Therefore, proactive strategies, including optimization of crop management,

investment in yield-enhancing technologies, and adaptive policy measures, are essential to safeguard future grain production and enhance national food security.

5. Discussions

This study proposes a dual-factor yield-prediction model that couples the one-month SPEI1 with rice-sown area, filling the gap left by previous analyses that treated climatic and land-use drivers in isolation. By integrating the two drivers, the model not only raises the accuracy of rice-yield forecasts but also supplies a theoretical basis for regional agricultural policy.

Results show that SPEI1 in September (SPEI1-Sep.) exerts the strongest negative effect on yield, especially during the grain-filling stage. Excess soil moisture reduces photosynthetic efficiency and suppresses growth, a finding consistent with Vicente-Serrano et al. [28]. The positive contribution of sown area is re-confirmed. Compared with climate variables, land-use policy influences sown area more directly and responds quickly to market demand. In Northeast China, rice area is largely shaped by national policies such as land-protection quotas and grain subsidies [11]. Consequently, land-use change is a non-negligible predictor of yield; future work should explore how policy intervention can optimize sown area and stabilize production. Model inter-comparison indicates that ARIMA significantly outperforms RF and XGBoost. ARIMA's strength in handling non-stationary time series allows it to capture the non-linear interplay between climate and land use, an advantage for agricultural forecasting that has been validated by Mishra et al. [33].

6. Conclusions

Against a backdrop of intensifying climate change and food-security risk, this paper constructs a dual-factor framework that merges SPEI and sown area to quantify the combined impacts of climate variability and land-use change on rice yield in the three Northeastern provinces.

At national level, SPEI1-Sep is the dominant yield modifier, and its predictive power is amplified in the Northeast because the region's cool climate and low effective accumulated temperature magnify water-balance effects [28]. Moderate drought elevates soil temperature and photosynthetic efficiency, accelerating grain filling, whereas excessive moisture hampers the process [50]. These insights provide a reference for climate-adaptive rice management. Relative to single-factor models, the dual-factor framework portrays yield trajectories under climate change more comprehensively and reflects inter-provincial agro-climatic heterogeneity. Among the tested algorithms, ARIMA delivers the lowest prediction error for the 1990–2022 period thanks to its capacity to deal with non-stationarity and to represent the dynamic climate–yield response. Forward-looking simulations for 2023–2027 suggest: Heilongjiang's output will undergo minor adjustments driven by sown-area fluctuations before stabilizing, Jilin will maintain mild growth under a wetter trend, demonstrating moderate resilience, and Liaoning will enter a dynamic-equilibrium phase after recent policy realignments. These provincial divergences reveal the complex coupling among natural constraints, policy interventions and market mechanisms, underscoring the need for province-specific food-security strategies.

Although the framework performs well in the Northeast, it still omits auxiliary climatic variables such as accumulated temperature and sunshine hours, limiting its representation of compound stresses. Future research should assimilate multi-source data, couple process-based crop models with machine learning, quantify interactions among water, heat and radiation, and conduct sensitivity and uncertainty analyses to enhance robustness. Cross-disciplinary and cross-institutional collaboration will be essential to meet the twin challenges of climate change and agricultural land-use evolution.

Author Contributions: Conceptualization, S.N. and Z.-Q.J.; methodology, S.N. and Z.-Q.J.; software, S.N.; validation, S.N. and Z.-Q.J.; formal analysis, Z.-Q.J.; investigation, S.N.; resources, S.N. and Z.-Q.J.; data curation, S.N. and Z.-Q.J.; writing—original draft preparation, S.N.; writing—review and editing, Z.-Q.J.; visualization, S.N. and Z.-Q.J.; supervision, Z.-Q.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partly supported by the National Natural Science Foundation of China (72171083).

Data Availability Statement: The dataset is publicly available at http://www.stats.gov.cn/sj/, http://www.moa.gov.cn/govpublic/ and https://data.tpdc.ac.cn/home, accessed on 1 November 2025.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT 4o and Grammarly for grammatical editing and enhancement. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

Abbreviations

The following abbreviations are used in this manuscript:

SPEI	standardized precipitation evapotranspiration index
RF	random forest
XGBoost	extreme gradient boosting
ARIMA	autoregressive integrated moving average
LSTM	long short-term memory networks
GRU	gated recurrent unit
PDSI	Palmer Drought Severity Index
SPI	standardized precipitation index

References

1. Lin, F.; Li, X.; Jia, N.; Feng, F.; Huang, H.; Huang, J.; Fan, S.; Ciais, P.; Song, X.P. The impact of Russia-Ukraine conflict on global food security. *Glob. Food Secur.* **2023**, *36*, 100661. [CrossRef]
2. Rudolfsen, I.; Bartusevic, H.; van Leeuwen, F.; Østby, G. War and food insecurity in Ukraine. *World Dev.* **2024**, *180*, 106647. [CrossRef]
3. Valera, H.G.A.; Mishra, A.K.; Pede, V.O.; Yamano, T.; Dawe, D. Domestic and international impacts of rice export restrictions: The recent case of Indian non-basmati rice. *Global Food Secur.* **2024**, *41*, 100754. [CrossRef]
4. Toromade, A.S.; Soyombo, D.A.; Kupa, E.; Ijomah, T.I. Reviewing the impact of climate change on global food security: Challenges and solutions. *Int. J. Appl. Res. Soc. Sci.* **2024**, *6*, 1403–1416. [CrossRef]
5. Shahbaz, K.; Muzaffar, M. The ripple effects of the Russia-Ukraine war on south Asian economies: Trade, energy, and inflation in a geopolitical crisis. *Pak. Soc. Sci. Rev.* **2025**, *9*, 20–35. [CrossRef]
6. Liu, Y.; Yang, L.; Zhang, J.; Cui, Q.; Liu, Y.; Nie, F.; Hu, Y. Aggravating effects of food export restrictions under climate change on food security: An analysis of rice economy based on alternative indicators. *Clim. Change Econ.* **2022**, *13*, 2240006. [CrossRef]
7. Cui, K.; Shoemaker, S.P. A look at food security in China. *Npj Sci. Food* **2018**, *2*, 4. [CrossRef]
8. Hou, D.; Chen, J.; Dong, J.; Ji, C.; Feng, J.; Du, G.; Yang, L. A 30-m annual paddy rice dataset in northeastern China during period 2000–2023. *Sci. Data* **2025**, *12*, 1355. [CrossRef]
9. Gao, J.; Faye, B.; Tian, R.; Du, G.; Zhang, R.; Biot, F. Understanding the impact of climatic events on optimizing agricultural production in northeast China. *Atmosphere* **2025**, *16*, 704. [CrossRef]
10. Meng, F.; Xie, K.; Liu, P.; Chen, H.; Wang, Y.; Shi, H. Extreme precipitation trends in northeast China based on a non-stationary generalized extreme value model. *Geosci. Lett.* **2024**, *11*, 13. [CrossRef]
11. Jin, Y.; Gardebroek, C.; Heerink, N. The impact of Chinese rice support policies on rice acreages. *Food Secur.* **2024**, *16*, 705–719. [CrossRef]
12. Heerink, N.; Qu, F.; Kuiper, M.; Shi, X.; Tan, S. Policy reforms, rice production and sustainable land use in China: A macro-micro analysis. *Agric. Syst.* **2007**, *94*, 784–800. [CrossRef]
13. Yi, F.; Sun, D.; Zhou, Y. Grain subsidy, liquidity constraints and food security-impact of the grain subsidy program on the grain-sown areas in China. *Food Policy* **2015**, *50*, 114–124. [CrossRef]

14. Hultgren, A.; Carleton, T.; Delgado, M.; Gergel, D.R.; Greenstone, M.; Houser, T.; Hsiang, S.; Jina, A.; Kopp, R.E.; Malevich, S.B.; et al. Impacts of climate change on global agriculture accounting for adaptation. *Nature* **2025**, *642*, 644–652. [CrossRef]
15. Prabnakorn, S.; Maskey, S.; Suryadi, F.X.; de Fraiture, C. Rice yield in response to climate trends and drought index in the mun river basin, Thailand. *Sci. Total Environ.* **2018**, *621*, 108–119. [CrossRef]
16. Sharma, R.K.; Kumar, S.; Vatta, K.; Bheemanahalli, R.; Dhillon, J.; Reddy, K.N. Impact of recent climate change on corn, rice, and wheat in southeastern USA. *Sci. Rep.* **2022**, *12*, 16928. [CrossRef]
17. Dai, A.; Trenberth, K.; Qian, T. A global dataset of palmer drought severity index for 1870–2002: Relationship with soil moisture and effects of surface warming. *J. Hydrometeorol.* **2004**, *5*, 1117–1130. [CrossRef]
18. Kumar, M.N.; Murthy, C.S.; Sai, M.V.R.S.; Roy, P.S. On the use of standardized precipitation index (SPI) for drought intensity assessment. *Meteorol. Appl.* **2009**, *16*, 381–389. [CrossRef]
19. Kogan, F. World droughts in the new millennium from AVHRR-based vegetation health indices. *Eos Trans. AGU* **2002**, *83*, 557–563. [CrossRef]
20. Wan, Z.; Wang, P.; Li, X. Using modis land surface temperature and normalized difference vegetation index products for monitoring drought in the southern great plains, USA. *Int. J. Remote Sens.* **2004**, *25*, 61–72. [CrossRef]
21. Wu, M.-x.; Lu, H.-q. A modified vegetation water supply index (MVWSI) and its application in drought monitoring over Sichuan and Chongqing, China. *J. Integr. Agric.* **2016**, *15*, 2132–2141. [CrossRef]
22. Wang, Q.; Wu, J.; Li, X.; Zhou, H.; Yang, J.; Geng, G.; An, X.; Liu, L.; Tang, Z. A comprehensively quantitative method of evaluating the impact of drought on crop yield using daily multi-scale SPEI and crop growth process model. *Int. J. Biometeorol.* **2017**, *61*, 685–699. [CrossRef]
23. Feng, K.; Su, X. Spatiotemporal characteristics of drought in the Heihe river basin based on the extreme-point symmetric mode decomposition method. *Int. J. Disaster Risk Sci.* **2019**, *10*, 591–603. [CrossRef]
24. Yu, M.; Li, Q.; Hayes, M.J.; Svoboda, M.D.; Heim, R.R. Are droughts becoming more frequent or severe in China based on the standardized precipitation evapotranspiration index: 1951–2010? *Int. J. Climatol.* **2014**, *34*, 545–558. [CrossRef]
25. Yang, P.; Xia, J.; Zhang, Y.; Zhan, C.; Qiao, Y. Comprehensive assessment of drought risk in the arid region of northwest China based on the global palmer drought severity index gridded data. *Sci. Total Environ.* **2018**, *627*, 951–962. [CrossRef]
26. Lesk, C.; Coffel, E.; Winter, J.; Ray, D.; Zscheischler, J.; Seneviratne, S.I.; Horton, R. Stronger temperature-moisture couplings exacerbate the impact of climate warming on global crop yields. *Nat. Food* **2021**, *2*, 683–691. [CrossRef]
27. Mondal, K.; Kar, R.K.; Chakraborty, A.; Dey, N. Concurrent effect of drought and heat stress in rice (*Oryza sativa* L.): Physio-biochemical and molecular approach. *3 Biotech* **2024**, *14*, 132. [CrossRef]
28. Vicente-Serrano, S.M.; Beguería, S.; López-Moreno, J.I. A multi-scalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *J. Clim.* **2010**, *23*, 1696–1718. [CrossRef]
29. Sohn, S.J.; Ahn, J.B.; Tam, C.Y. Six month-lead downscaling prediction of winter to spring drought in South Korea based on a multi-model ensemble. *Geophys. Res. Lett.* **2013**, *40*, 579–583. [CrossRef]
30. Xu, L.; Chen, N.; Yang, C.; Zhang, C.; Yu, H. A parametric multivariate drought index for drought monitoring and assessment under climate change. *Agric. For. Meteorol.* **2021**, *310*, 108657. [CrossRef]
31. Han, J.; Singh, V.P. A review of widely used drought indices and the challenges of drought assessment under climate change. *Environ. Monit. Assess.* **2023**, *195*, 1438. [CrossRef]
32. Everingham, Y.; Sexton, J.; Skocaj, D.; Inman-Bamber, G. Accurate prediction of sugarcane yield using a random forest algorithm. *Agron. Sustain. Dev.* **2016**, *36*, 27. [CrossRef]
33. Mishra, P.; Al Khatib, A.M.G.; Yadav, S.; Ray, S.; Lama, A.; Kumari, B.; Sharma, D.; Yadav, R. Modeling and forecasting rainfall patterns in India: A time series analysis with XGBoost algorithm. *Environ. Earth Sci.* **2024**, *83*, 163. [CrossRef]
34. Haider, S.A.; Naqvi, S.R.; Akram, T.; Umar, G.A.; Shahzad, A.; Sial, M.R.; Khaliq, S.; Kamran, M. LSTM neural network based forecasting model for wheat production in Pakistan. *Agronomy* **2019**, *9*, 72. [CrossRef]
35. Leukel, J.; Zimpel, T.; Stumpe, C. Machine learning technology for early prediction of grain yield at the field scale: A systematic review. *Comput. Electron. Agric.* **2023**, *207*, 107721. [CrossRef]
36. Li, Y.; Yuan, X.; Zhang, H.; Wang, R.; Wang, C.; Meng, X.; Zhang, Z.; Wang, S.; Yang, Y.; Han, B.; et al. Mechanisms and early warning of drought disasters: Experimental drought meteorology research over China. *Bull. Am. Meteorol. Soc.* **2019**, *100*, 673–687. [CrossRef]
37. Wang, Y.; Zhao, W.; Zhang, Q.; Yao, Y.b. Characteristics of drought vulnerability for maize in the eastern part of Northwest China. *Sci. Rep.* **2019**, *9*, 964. [CrossRef]
38. Thom, H.C. A note on the gamma distribution. *Mon. Weather Rev.* **1958**, *86*, 117–122. [CrossRef]
39. Zhang, Q.; Miao, C.; Su, J.; Gou, J.; Hu, J.; Zhao, X.; Xu, Y. A new high-resolution multi-drought-index dataset for mainland China. *Earth Syst. Sci. Data* **2025**, *17*, 837–853. [CrossRef]

40. Kamruzzaman, M.; Almazroui, M.; Salam, M.A.; Mondol, M.A.H.; Rahman, M.M.; Deb, L.; Kundu, P.K.; Zaman, M.A.U.; Islam, A.R.M.T. Spatiotemporal drought analysis in Bangladesh using the standardized precipitation index (SPI) and standardized precipitation evapotranspiration index (SPEI). *Sci. Rep.* **2022**, *12*, 20694. [CrossRef]
41. Mendicino, G.; Senatore, A. Regionalization of the Hargreaves coefficient for the assessment of distributed reference evapotranspiration in southern Italy. *J. Irrig. Drain. Eng.* **2013**, *139*, 349–362. [CrossRef]
42. Wang, Q.; Zeng, J.; Qi, J.; Zhang, X.; Zeng, Y.; Shui, W.; Xu, Z.; Zhang, R.; Wu, X.; Cong, J. A multi-scale daily SPEI dataset for drought characterization at observation stations over mainland China from 1961 to 2018. *Earth Syst. Sci. Data* **2021**, *13*, 331–341. [CrossRef]
43. Amnuaylojaroen, T.; Chanvichit, P. Historical analysis of the effects of drought on rice and maize yields in Southeast Asia. *Resources* **2024**, *13*, 44. [CrossRef]
44. Nie, T.; Liu, X.; Chen, P.; Jiang, L.; Sun, Z.; Yin, S.; Wang, T.; Li, T.; Du, C. Characterizing droughts during the rice growth period in northeast China based on daily spei under climate change. *Plants* **2025**, *14*, 30. [CrossRef]
45. Sein, Z.M.M.; Zhi, X.; Ogou, F.K.; Nooni, I.K.; Sian, K.T.C.L.K.; Gnitou, G.T. Spatio-temporal analysis of drought variability in myanmar based on the standardized precipitation evapotranspiration index (spei) and its impact on crop production. *Agronomy* **2021**, *11*, 1691. [CrossRef]
46. Ho, S.L.; Xie, M. The use of ARIMA models for reliability forecasting and analysis. *Comput. Ind. Eng.* **1998**, *35*, 213–216. [CrossRef]
47. Pang, R.; Sun, D.; Sun, W. Spatiotemporal Variations in Grain Yields and Their Responses to Climatic Factors in Northeast China During 1993–2022. *Land* **2025**, *14*, 1693. [CrossRef]
48. Li, J.; Xie, Y.; Liu, L.; Song, K.; Zhu, B. Long short-term memory neural network with attention mechanism for rice yield early estimation in qian gorlos county, northeast China. *Agriculture* **2025**, *15*, 231. [CrossRef]
49. Sbai, Z. Deep learning models and their ensembles for robust agricultural yield prediction in saudi arabia. *Sustainability* **2025**, *17*, 5807. [CrossRef]
50. Zhang, H.; Li, H.; Yuan, L.; Wang, Z.; Yang, J.; Zhang, J. Post-anthesis alternate wetting and moderate soil drying enhances activities of key enzymes in sucrose-to-starch conversion in inferior spikelets of rice. *J. Exp. Bot.* **2012**, *63*, 215–227. [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.