earth

MDPI

Article

Machine Learning-Based Regional Assessment of Second-Crop Corn Yield Using Climatic and Water-Balance Variables

Guilherme B. Torsoni [1], Lucas E. O. Aparecido [2] and Fernando G. Macedo [1,*]

[1] Instituto Federal de Mato Grosso do Sul, Naviraí 79947-334, Brazil; guilherme.torsoni@ifms.edu.br
[2] Instituto Federal do Sul de Minas Gerais, Muzambinho 37890-000, Brazil; lucas.aparecido@ifsuldeminas.edu.br
* Correspondence: fernando.macedo@ifms.edu.br; Tel.: +55-67-2020-6600

Abstract

Second-crop corn is an important component of Brazilian grain production and is highly exposed to climatic variability because crop development occurs during the transition from the rainy to the drier season. This study assessed the performance of machine learning algorithms for estimating second-crop corn yield within the 2012–2023 historical data domain in Mato Grosso do Sul, Brazil. Municipal yield data from 22 municipalities and daily climatic data covering the period from 2012 to 2023 were analyzed. Yield data were obtained from Aprosoja/MS, whereas climatic variables, including solar radiation, air temperature, evapotranspiration, relative humidity, wind speed, and precipitation, were obtained from NASA/POWER. Pearson correlation analysis was used as an initial screening procedure to identify climatic variables associated with yield variability. Multiple linear regression (MLR), a multilayer perceptron (MLP), a support vector machine (SVM), Random Forest (RF), XGBoost, and a gradient boosting regressor (GBR) were evaluated. Under the observation-level random partition adopted in this study, RF and XGBoost showed the highest performance in the randomly held-out testing subset, both reaching R^2 values of approximately 0.80. However, XGBoost exhibited near-perfect calibration performance followed by a substantial decrease during testing (MAPE = 14.05%; R^2 = 0.80), indicating overfitting. Random Forest achieved similar testing performance with a smaller calibration-to-testing gap. Therefore, under the specific random partition used in this study, RF was considered the preferred model. In contrast, GBR failed to provide reliable predictive performance under the present dataset and configuration, with a testing MAPE of 547.78%, and was therefore considered too unreliable for interpretation in this study. Importantly, because the partition did not preserve temporal or spatial independence, these results represent generalization within the 2012–2023 historical data domain and should not be interpreted as evidence of future-year or spatially independent predictive performance. Because the dataset was partitioned randomly at the observation level without preserving temporal independence, the reported performance represents generalization within the 2012–2023 historical data domain and should not be interpreted as future-year forecasting performance.

Keywords: machine learning; climate data; big data; agricultural planning

Check for updates

Academic Editors: Charles Jones, Dimitrios E. Tsesmelis, Pantelis E. Barouchas, Georgios Bourantas and Kleomenis Kalogeropoulos

Received: 18 August 2026
Revised: 13 October 2026
Accepted: 20 October 2026
Published: 23 October 2026

Copyright: © 2026 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.

1. Introduction

Corn has been integral to human culture for thousands of years [1]. An essential crop across various regions of North and South America, it eventually made its way to Europe

and other global territories [2]. Currently, corn is one of the most cultivated crops in the world, with an estimated production of more than 1 billion tons annually [3]. The use of corn for ethanol (bioenergy) production is the fastest growing application worldwide. Brazil is one of the largest corn producers in the world, with a production of 114.7 million tons of corn for the 2025/2026 growing season [4]. The crop is grown in different regions of Brazil, with the highest production levels in the South and Midwest regions (76%) [5,6]. Brazil became the world's largest corn exporter in the agricultural year that ended in August 2023, according to information from the United States Department of Agriculture [7]. Brazil was responsible for 32% of global corn exports, with shipments of 56 million tons, surpassing the United States, which led this market for more than a century [8].

Agricultural cultivation in some parts of Brazil differs from the rest of the world due to the ability to produce two consecutive crops without the need for irrigation [9–11]. In these areas, soybeans are generally planted in October/November, with harvesting in February/March. Following this (i.e., as the second crop), corn is planted in February/March and harvested in June. The State of Mato Grosso do Sul has a land area similar to that of Germany; therefore, it is larger than the United Kingdom, Italy, and Poland, and is nearly four times the size of Portugal. Mato Grosso do Sul is one of the main states producing second-crop corn in Brazil. Although cultivating two crops consecutively is highly advantageous, the water regime of the second crop is lower than that of the first crop, making the second crop more vulnerable to climatic factors that can impact agricultural production [11–13].

Thus, the climate has a strong impact on agricultural activities [14–16], consisting of the main factors responsible for fluctuations and oscillations in grain yield of diverse crops [15,17]. The critical meteorological elements in agricultural production are air temperature, solar radiation, and precipitation [18].

Agrometeorological models have become increasingly important in modern agriculture, helping farmers make informed decisions about crop management and agricultural practices in relation to climate [19,20]. These models use meteorological data to provide valuable information about the behavior and yield of agricultural systems [21]. One of the main benefits of agrometeorological models is their ability to predict how crops and agricultural systems will respond to climate conditions [22,23]. These models simulate crop growth under different environmental scenarios, supporting decision-making regarding planting schedules, fertilization, and management practices [24,25]. Water-balance models also contribute to optimizing irrigation management [26–28].

Climate change represents a significant challenge for agriculture, affecting crop yields and requiring adaptation strategies to ensure food security and sustainability [29]. In this context, the development of advanced tools for agricultural zoning has become essential [30]. These approaches help identify suitable cultivation areas, predict risks, and support decision-making processes [31].

The incorporation of artificial intelligence (AI) into agricultural systems represents a major technological advance [32–34]. These approaches enable the analysis of large climate datasets and the identification of complex patterns not detectable by traditional methods [35].

Recent advances in artificial intelligence have significantly improved agricultural monitoring and decision-making processes, particularly through AI-driven crop monitoring systems capable of optimizing resource allocation, temporal crop assessment, and sustainable agricultural management. Recent studies have demonstrated the strong potential of machine learning and deep learning approaches combined with multi-date aerial imagery, transfer learning, and automatic annotation workflows for automated crop growth monitoring and precision agriculture applications. These technologies enable the continuous

evaluation of crop development under different environmental conditions, supporting more accurate and data-driven agricultural management strategies while improving the efficiency and sustainability of crop production systems.

Machine learning techniques have demonstrated high efficiency in predicting crop yields under varying climate conditions [36–38], allowing improvements in agricultural planning and management [39,40].

Although machine learning approaches have been widely applied to crop-yield prediction, comparatively few studies have focused specifically on second-crop corn under the production conditions of Central–Western Brazil, where crop development occurs during the transition from the rainy to the drier season and is strongly influenced by thermal and water variability. In this context, the contribution of the present study lies in the integration of multi-year municipal yield records with daily climatic and water-balance variables across 22 municipalities over a 12-year period, together with the comparative evaluation of six machine learning approaches and the assessment of spatial agreement between observed and predicted yields. This integrated regional framework provides a regional assessment of the performance and limitations of established machine learning methods for estimating second-crop corn yield within the 2012–2023 historical data domain in Mato Grosso do Sul.

Accordingly, this study represents a regional assessment of established machine learning methods under the stated data and validation limitations; it does not introduce a new algorithm and should not be interpreted as a validated framework for future-year or operational yield forecasting.

2. Material and Methods

2.1. Big Data Collection

The study used a historical series of climate and phytotechnical data from corn production areas in the State of Mato Grosso do Sul (MS), one of the main grain-producing states in Brazil. The analysis was conducted for the following municipalities: Amambai (Ama), Aral Moreira (Ara), Bandeirantes (Ban), Caarapó (Caa), Camapuã (Cam), Campo Grande (CaG), Chapadão do Sul (Cha), Costa Rica (Cos), Coxim (Cox), Douradina (Dou), Dourados (DoD), Itaporã (Ita), Itaquiraí (ItQ), Laguna Carapã (Lag), Maracaju (Mar), Naviraí (Nav), Nova Alvorada do Sul (Nov), Ponta Porã (Pon), Rio Brilhante (Rio), São Gabriel do Oeste (São), Sidrolândia (Sid), and Sonora (Son). Yield data were collected at the municipal scale, representing aggregated productivity values for key corn-producing cities, and were obtained from the Association of Soybean Producers of Mato Grosso do Sul (Aprosoja/MS—Campo Grande, MS, Brazil), ensuring data reliability and representativeness. Figure 1 shows the locations in Mato Grosso do Sul that are most representative of corn production.

Yield data (kg ha^{-1}) were meticulously compiled over an extensive period from 2012 to 2023. The 2012–2023 period was selected because it represents the longest interval for which consistent and simultaneous municipal yield and climate data were available for all studied locations. Although a longer time series would be desirable for long-term climatological characterization, this 12-year dataset encompasses substantial interannual variability in both climatic conditions and corn yield and was therefore considered adequate for the regional model assessment conducted within the 2012–2023 historical data domain. It is important to highlight that the yield dataset inherently incorporates the effects of technological advances over time, including genetic improvement and changes in crop management practices, such as fertilization, pest control, and planting strategies. As the data represent observed yields at the municipal level over a 12-year period, these factors are implicitly included in the yield response. Therefore, the modeling approach does not isolate these effects individually, but rather captures their combined influence within the

historical yield variability, allowing the models to be as comprehensive as possible without restricting the analysis to specific technological conditions.

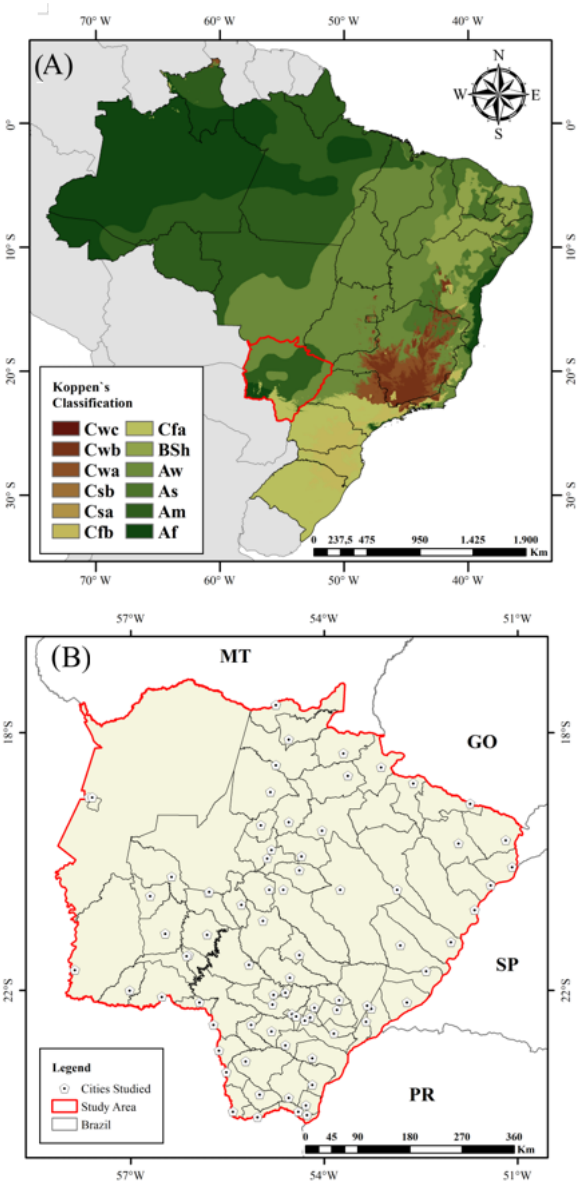


Figure 1. (**A**) Köppen climate classification map of Brazil, highlighting the State of Mato Grosso do Sul (MS) within the national territory. (**B**) Geographic location of the study area in Mato Grosso do Sul, Brazil, showing the municipalities analyzed (points) and the state boundary (red line). Neighboring states are also indicated for spatial reference.

These comprehensive data were acquired through the Association of Soybean Producers of Mato Grosso do Sul (Aprosoja/MS), thus ensuring the quality and reliability of information. The temporal range of this collection allows for a robust and comprehensive analysis of corn production trends in these representative locations of the state. This detailed database provides a basis for assessing historical corn production variability and its associations with climatic conditions within the 2012–2023 study period.

Data on global solar radiation (Qg), mean (Tm), maximum (Tmax), and minimum (Tmin) air temperature (°C), evapotranspiration (ETP), relative humidity (RH, %), wind speed (U2), and precipitation (P, mm) for the period 2012–2023 were obtained from the National Aeronautics and Space Administration/Prediction of Worldwide Energy Resources (NASA/POWER) platform on a daily scale. This data platform was developed to provide

meteorological information, and data was compiled from multiple sources and derived from grids with a spatial resolution of 0.25° (latitude–longitude).

NASA/POWER provides analysis-ready climate data derived from a combination of satellite observations and model-based products. Historical surface meteorological variables are primarily based on NASA reanalysis/model products, including MERRA-2, which assimilate observational information, while solar-radiation variables are derived from satellite-based products such as CERES and SRB. Thus, the POWER dataset used here should not be interpreted as presenting direct in situ observations, but rather as presenting gridded satellite- and model/reanalysis-derived climate products.

For each municipality, NASA/POWER climatic data were extracted using a representative geographic coordinate located within the municipal area. The climatic values associated with the corresponding 0.25° × 0.25° grid cell were therefore assigned to that municipality throughout the study period. No area-weighted averaging among multiple grid cells was performed. This approach provided a consistent climatic time series for each municipality, although the spatial resolution of NASA/POWER may not capture all local-scale climatic variability within municipal boundaries.

Potential evapotranspiration (ETP) was calculated from the obtained climate data using the Penman–Monteith (PM) method [41]. The PM method is presented in Equation (1):

$$ETP = \frac{0.408 \times s \times (Rn - G) + \frac{\gamma \times 900 \times U_2 \times (es - ea)}{T + 273}}{s + \gamma \times (1 + 0.34 \times U_2)} \quad (1)$$

where ETP is the potential evapotranspiration, in mm day^{-1}; s is the slope of the vapor saturation curve, in KPa $°C^{-1}$; U_2 is the mean daily wind speed at 2 m high, in m s^{-1}; Rn is the net radiation, in MJ m^{-2} day^{-1}; G is the heat flux density, in MJ m^{-2} day^{-1}; T is the mean daily air temperature, in °C; γ is the psychometric constant, equal to 0.063 KPa $°C^{-1}$; *ea* is the partial vapor pressure, in KPa; and *es* is the vapor saturation pressure, daily mean, in Kpa.

The climatological water balance was generated for all studied locations according to the methodology of Thornthwaite and Mather [42] (Equations (2)–(7)). A standardized available water capacity (AWC) of 40 mm [43,44] was adopted uniformly for all municipalities to calculate the climatological water balance. This value represents a simplified assumption of the amount of plant-available water stored within the effective rooting zone and was not adjusted according to municipality-specific soil texture or soil hydraulic properties [2–7].

Therefore, the derived soil water storage (STO), water deficit (DEF), and water surplus (EXC) should be interpreted as relative indicators within the adopted simplified water-balance framework rather than as municipality-specific estimates of actual soil water availability.

$$If\ (P - PET)_i < 0 = \begin{cases} NAC_i = NAC_{i-1} + (P - PET)_i \\ STO_i = AWC \times e^{\frac{(NAC_i)}{AWC}} \end{cases} \quad (2)$$

$$If\ (P - PET)_i \geq 0 = \begin{cases} STO_i = (P - PET)_i + STO_{i-1} \\ NAC_i = AWC \times ln^{\frac{(STO_i)}{AWC}} \end{cases} \quad (3)$$

$$ALT_i = STO_i - STO_{i-1} \quad (4)$$

$$ETR_i = \begin{cases} P + |ALT_i| & , if\ ALT < 0 \\ ETP_i & , if\ ALT \geq 0 \end{cases} \quad (5)$$

$$DEF_i = ETP_i - ETR_i \quad (6)$$

$$EXC_i = \begin{cases} 0 & , \text{if } ARM_i < AWC \\ (P - PET)_i - ALT_i & , \text{if } ARM_i = AWC \end{cases} \quad (7)$$

where AWC is the available water capacity (mm); STO is the soil water storage (mm); EXC is the water surplus in the soil–plant–atmosphere system (mm); DEF is the water deficit of the soil–plant–atmosphere system (mm); NAC is the negative (P-ETo) value; P is the precipitation (mm); ETP is the potential evapotranspiration (mm); ETR is the real evapotranspiration (mm); ALT is the alteration in soil water storage of the current month and the previous month (mm); and i is a given month.

Pearson's univariate correlation analysis was initially applied to evaluate the relationship between climatic variables and second-crop corn yield at the municipal level. Based on this analysis, the climatic variables showing the strongest correlations were selected as independent variables (features) in the machine learning models, while corn yield was defined as the dependent variable. This approach allowed the integration of climate and production data by linking climatic variables showing the strongest correlations with yield variability [45].

Pearson's correlation coefficients were calculated between each daily climatic variable and corn yield throughout the March–June growing period. The variables showing the strongest absolute correlations with yield were prioritized as predictors. Pearson correlation analysis was also used as a preliminary screening tool to identify potential redundancy among climatic variables; however, no formal correlation threshold, variance inflation factor (VIF), or condition-number criterion was applied.

Pearson correlation was used only as a preliminary univariate screening criterion and, therefore, may not capture nonlinear relationships or higher-order interactions among climatic predictors. Recursive feature elimination (RFE), Random Forest feature importance, variance inflation factor (VIF), and condition-number analyses were not performed in the present study. Consequently, the selected predictor set should not be interpreted as an optimized feature subset, but rather as a reduced set of climatically relevant variables identified through correlation screening and subsequent model calibration.

The independent variables (FEATURES) were the predominant daily climatic variables within the March–June crop-development window of second-crop corn: global solar radiation (Qg), mean (Tm), maximum (Tmax), and minimum (Tmin) air temperature (°C), evapotranspiration (ETP), relative humidity (RH, %), wind speed (U2), precipitation (P, mm), water deficit (mm), and water surplus (mm). Importantly, climate data from March to June were used, as this corresponds to the average period from planting to the beginning of maturation of corn plants in the State of Mato Grosso do Sul. The dependent variable of the model was the corn yield (kg ha^{-1}).

Potential redundancy among climatic predictors was considered during the preliminary Pearson correlation screening used for feature selection. This procedure was intended to reduce the inclusion of variables carrying overlapping climatic information. However, no additional formal multicollinearity diagnostic, such as the variance inflation factor (VIF), was calculated, and residual redundancy among predictors cannot be completely excluded.

The complete dataset comprised 264 municipality-year observations (22 municipalities × 12 years). For the MLR model, the number of predictors was treated as a tuning parameter, and the best-performing configuration retained 14 features. Although this number is moderate relative to the total dataset size, the effective calibration sample was smaller because 70% of the observations were used for model fitting. Therefore, the possibility of model instability and redundancy among predictors cannot be completely excluded, particularly because several climatic variables are physically related.

In this study, ETP represents the atmospheric water demand estimated by the Penman–Monteith method and is expressed in mm day^{-1}. The derived water-balance

variables were expressed in millimeters: DEF represents the amount by which atmospheric water demand exceeds the water effectively available to the crop, whereas EXC represents water remaining after soil water storage reaches the adopted available water capacity of 40 mm. STO represents soil water storage, and ETR represents actual evapotranspiration. These variables were calculated from daily meteorological series and aligned with the second-crop corn growing period considered in this study, from March to June.

2.2. Calibration and Testing of Artificial Intelligence

The machine learning models used at this stage were multiple linear regression (MLR), a multilayer perceptron (MLP), a support vector machine (SVM), Random Forest (RF), extreme gradient boosting (XGBOOST), and a gradient boosting regressor (GBR) [46,47]. The complete dataset comprised 264 municipality-year observations, which were randomly divided at the observation level, with 70% assigned to the calibration subset and 30% assigned to a randomly held-out testing subset [48]. A fixed random seed was used to ensure reproducibility. The random partition was performed regardless of the year and municipality of observation; therefore, the procedure did not preserve temporal or spatial independence between the calibration and testing subsets. Consequently, observations from the same historical period and from municipalities represented in the calibration dataset could also occur in the testing subset. The randomly held-out testing subset should therefore be interpreted as an internal assessment of model generalization within the 2012–2023 historical data domain, rather than as temporally or spatially independent validation.

Consequently, there were no distinct training-year and testing-year ranges; observations from the 2012–2023 period could be represented in both subsets according to the random partition.

A fixed random seed was used to ensure reproducibility of the partitioning procedure; because the partitioning was performed randomly at the observation level, chronological and spatial independence were not preserved. Thus, the randomly held-out testing subset evaluates model performance for unseen municipality-year observations drawn from the same 2012–2023 historical domain, rather than true temporal extrapolation to future growing seasons or spatial extrapolation to previously unseen municipalities. The reported testing metrics may therefore be optimistic relative to those expected under fully temporally or spatially independent validation. Accordingly, the reported R^2, MAPE, and other testing metrics should not be interpreted as estimates of operational future-year forecasting performance.

Climatic predictors corresponding to the March–June crop-development period of each municipality-year observation were used to estimate yield for that same crop season. Therefore, the present framework represents within-season yield prediction within the historical dataset rather than pre-season forecasting of future years. Leave-One-Year-Out, temporally ordered hold-out, rolling-window, or spatial cross-validation would be required to quantify model performance under temporal or spatial extrapolation, but these validation schemes were not implemented in the present study.

Hyperparameter selection was performed during the calibration stage by evaluating the predefined candidate values reported in Table 1, while the randomly held-out testing subset was not used for model fitting. A fixed random seed was maintained to ensure reproducibility. No learning-curve analysis or additional early-stopping procedure was performed; therefore, overfitting was assessed by comparing calibration and randomly held-out testing subset performance.

Table 1. Parameters defined and tested in each machine learning algorithm used to estimate second-crop corn yield within the 2012–2023 historical data domain in Mato Grosso do Sul.

Parameter	Tested Value
	MLR
k features	2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 15
cross-validation	2; 3; 4; 5; 6; 7; 8; 9
	MLP
hidden layer sizes	(10; 11; 12); (9; 10; 9); (7; 8; 10); (10; 10; 10); (25;11;7;5;3); (10; 10; 10); (20; 20; 20)
max iter	100; 1000; 2000; 3000;5000; 7000
cross-validation	2; 3; 4; 10; 15; 20
learning rate init	0.01; 0.05; 0.1; 0.5; 1; 5; 10
learning rate init	constant; adaptive
activation	relu; tanh; logistic
solver	lbfgs; adam
alpha	0.01; 0.05; 0.1; 0.5; 1; 5; 10
random state	10
	SVM
kernel	rbf; linear; poly; sigmoid
C	1; 10; 50; 100; 1000; 10,000
epsilon	0.1; 1; 4; 5; 10; 100
gamma	auto; scale
degree	1; 2; 3; 4; 7
coef0	0; 1
	RF
max depth	none; 10; 15; 30; 300
n estimators	10; 100; 200; 500; 1000; 10,000
	XGBOOST
max depth	1; 10; 15; 100
alpha	10; 20; 200; 2000
n estimators	300; 3000; 30,000
learning rate	0.01; 0.05; 0.1; 0.5; 0.2
colsample_bytree	0.01; 0.05; 0.1; 0.5; 0.3
	GBR
n estimators	10; 100; 500; 1000
max depth values	none; 5; 10; 100
random state values	10; 20; 30; 40; 50

C is a hyperparameter that controls the trade-off between the correct classification of training points and the maximization of the decision margin.

No additional independent validation subset or nested cross-validation procedure was used. Hyperparameter selection was conducted exclusively within the calibration dataset, using the predefined candidate values and internal cross-validation settings reported in Table 1, where applicable. The randomly held-out testing subset was kept completely separate from hyperparameter selection and model fitting and was used only for the final assessment of predictive performance.

All algorithms were evaluated using the same calibration/testing partition and the same candidate predictor pool obtained after the initial feature-screening procedure. However, the hyperparameter search spaces were necessarily model-specific because the algorithms had different structures and tuning parameters, as reported in Table 1. For MLR, the number of retained predictors (k_features) was additionally treated as a tuning parame-

ter [48]. Therefore, the comparison was based on a common dataset and predictor pool, but not on identical hyperparameter spaces across algorithms.

The second-crop corn planting period, from March to June [49], was considered to define the input climate variables. Accordingly, only climatic and water-balance variables corresponding to the March–June crop-development window were considered in the predictor-selection procedure. This temporal restriction was adopted to align the environmental predictors with the period encompassing the main vegetative and reproductive development of second-crop corn in Mato Grosso do Sul. The climatic predictors were organized as temporally indexed daily observations within the March–June crop-development window. The notation T—n was used only as a temporal index for air-temperature observations, with analogous indexing applied to P, Qg, Rn, RH, STO, DEF, and EXC. These indices should not be interpreted as representing specific phenological stages. Because exact municipality- and year-specific phenological dates were not available, the climatic variables were aligned to the broader March–June development period, which encompasses the main vegetative and reproductive phases of second-crop corn in Mato Grosso do Sul.

Throughout the manuscript, calibration refers to the subset used to fit the models, whereas randomly held-out testing refers to the observation-level subset used to assess generalization within the 2012–2023 historical data domain. The testing subset does not represent temporally independent validation.

2.3. Assessment of the Performance of AI Algorithms

The degree of learning was verified by the mean squared error (MSE), while the performance of the ML models in the randomly held-out testing subset was evaluated using mean absolute percentage error (MAPE), mean absolute error (MAE), and the adjusted coefficient of determination (R^2adj), as proposed by Mas and Ahlfeld [50], Ahangar et al. [51], and refs. [9–13], where YOBS is the observed yield (kg ha^{-1}), YEST is the estimated yield (kg ha^{-1}), N is the number of observations, k is the number of predictors, and $\bar{Y}$ is the mean observed yield.

After selecting the best-performing algorithm under the adopted evaluation framework, the observed and model-estimated yields, together with the differences between them, were interpolated using the Kriging method [52].

The Kriging interpolation was used primarily to provide an exploratory visualization of regional spatial patterns based on the municipal observations. Formal geostatistical cross-validation of the interpolated surfaces was not conducted; therefore, the resulting maps should not be interpreted as independently validated continuous yield surfaces. Rather, they should be interpreted only as exploratory spatial representations of the observed and model-estimated municipal-scale values.

This approach allowed the creation of maps that revealed the spatial variation in corn yields. These maps were generated by the QGIS 3.24 software, providing a regional visualization of spatial yield patterns. This spatially explicit approach is crucial to better understanding corn production patterns and can provide valuable insights for optimizing agricultural practices and strategic decision-making.

3. Results and Discussion

The analysis of the evolution of corn production in several municipalities in the State of Mato Grosso do Sul over time (Figure 2) showed a striking variation that highlights the effect of multiple factors on agricultural activity. Costa Rica and Chapadão do Sul stood out in 2023, reaching yields of 8635 kg ha^{-1} and 8636 kg ha^{-1}, respectively.

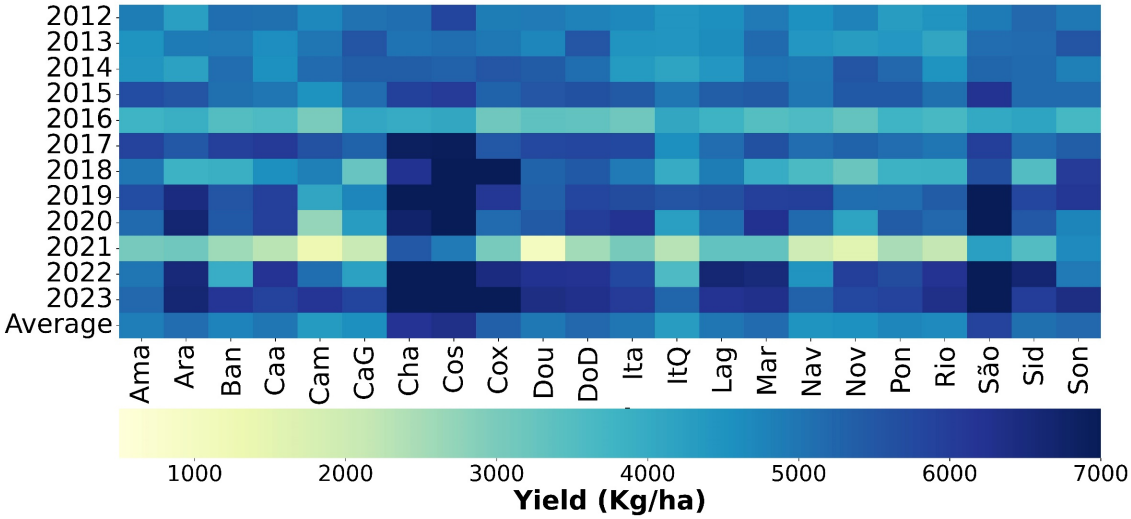


Figure 2. Heat map of temporal variation in corn yield for producing locations in Mato Grosso do Sul from 2012 to 2023. Legend: Amambai (Ama), Aral Moreira (Ara), Bandeirantes (Ban), Caarapó (Caa), Camapuã (Cam), Campo Grande (CaG), Chapadão do Sul (Cha), Costa Rica (Cos), Coxim (Cox), Douradina (Dou), Dourados (DoD), Itaporã (Ita), Itaquiraí (ItQ), Laguna Carapã (Lag), Maracaju (Mar), Naviraí (Nav), Nova Alvorada do Sul (Nov), Ponta Porã (Pon), Rio Brilhante (Rio), São Gabriel do Oeste (São), Sidrolândia (Sid), and Sonora (Son).

In a contrasting scenario, Douradina presented the lowest yield observed in the 2021 study, recording a production of only 986 kg ha^{-1}. This notably low performance is mainly due to the severe water shortages that affected the region throughout the year. The low yield observed during this period was associated with insufficient rainfall and elevated temperatures [53]. This scenario shows the urgent need to adopt more efficient and sustainable water resource management practices, aiming to minimize the devastating effects of water scarcity on agricultural yield.

The Amambai region experienced a significant fluctuation in its agricultural production in the period between 2012 and 2019. Initially, production in 2012 was 4883 kg ha^{-1}, but a drop was observed, reaching 3798 kg ha^{-1} in 2016. However, Amambai managed to reverse this negative trend in 2017, increasing its production to an impressive 5858.44 kg ha^{-1}.

On the other hand, Aral Moreira presented a different scenario. This region had a trajectory of sustained growth, starting from a production of 4244 kg ha^{-1} in 2012 and reaching a peak of 6648 kg ha^{-1} in 2020. This increase can be explained by a phenomenon related to the management of water resources and mild temperatures, essential for agricultural success in the region [54,55].

Bandeirantes recorded a remarkable corn yield, reaching 5938 kg ha^{-1} in 2017. This exemplary performance, however, faced a significant decline in subsequent years, culminating in a considerably lower production of 2584 kg ha^{-1} in 2021. This drastic reduction in yield coincided with adverse climatic conditions, including periods of prolonged drought [56,57].

Campo Grande, the capital of the State of Mato Grosso do Sul, showed a declining trend in agricultural production after 2017, evidenced by a decrease in yield, which fell from 5286 kg ha^{-1} to 2077 kg ha^{-1} in 2021 [58]. This period of low production coincided with adverse weather conditions. However, a notable turnaround was observed in 2023, when agricultural yield showed signs of significant recovery, reaching 5852.40 kg ha^{-1}.

The state faced a significant drought in 2016 and more significantly in 2021, affecting several areas, including agriculture. The lack of rain and adverse weather conditions led to a drastic reduction in soil moisture levels, negatively impacting agricultural yield. Corn is one of the state's main crops and suffered severely from these conditions. Water shortages

not only affected the germination and growth of corn plants but also limited farmers' ability to irrigate the crops adequately. As a result, corn production throughout the state experienced a sharp drop, compromising farmers' income and the availability of corn for the domestic market and exports [59,60].

The mean air temperature varies between 21 °C and 26 °C in the highest areas of the State of Mato Grosso do Sul (Figure 3A), with the Pantanal and the Northeast presenting higher temperatures [61]. This thermal variation has direct implications for corn physiology. Temperatures within this range can be ideal for corn growth, but higher temperatures, especially above 30 °C, may lead to heat stress. This stress may reduce photosynthesis, negatively affect grain formation, and reduce plant water use efficiency.

Precipitation in the state, with annual values that can reach 1689 mm in the South, Center, and North regions contrasting with up to 966 mm annually in the East and West (Figure 3B), plays a fundamental role in water availability for corn [62]. Water is essential for seed germination, plant growth, and the grain-filling stage, which are critical for determining the final yield. Regions with lower precipitation may face challenges related to water deficit, adversely affecting plant growth and potentially leading to the need for irrigation to maintain yield.

Furthermore, the distribution of precipitation throughout the growing cycle is as important as the total amount. Excessive rainfall during sowing or harvesting can cause soil erosion, loss of nutrients, and difficulties in operating agricultural machinery, while inadequate distribution of rainfall during the cycle can result in water stress during sensitive periods of crop development.

Corn irrigation during periods of water scarcity is a crucial agronomic strategy to maintain and potentially increase yield in regions affected by adverse climate variations. This practice can compensate for rainfall deficiency, ensuring the water supply necessary for crop development during the March–June growing window.

The evapotranspiration map (Figure 3C) shows values that reached 1820 mm annually in the Northeast and far East of the state, while other regions had approximately 940 mm. This may indicate substantial differences in crop water demand or environmental water availability between these areas.

Water storage (Figure 3D) is lowest in the Northwest and East of the state, varying between 33 and 0 mm annually. In contrast, the South and North boast higher storage capacity, with values of up to 59 mm annually. Thus, considerable spatial variability can be observed in water storage potential, which must be considered in the management of water resources.

There is little or no evidence of a water deficit (Figure 3E) in the Center or South of MS. However, the deficit is accentuated in the East and Northeast of the state, varying between 265 and 333 mm annually. The distribution of water deficits suggests potential areas for agricultural stress or the need for supplemental irrigation.

Moreover, water surplus (Figure 3F) shows higher potential in the South and Center of the state, with values that can reach up to 704 mm annually. This signals that these regions may have better conditions for recharging aquifers and supporting aquatic ecosystems, as well as indicating potential for agricultural practices that depend on higher water availability.

Negative correlations, such as that observed between mean air temperature in March (−0.558) (Figure 4A) and yield, indicate that higher temperatures during this month were associated with lower corn yield. Because exact municipality- and year-specific phenological dates were not available, this relationship should be interpreted as a climatic association within the March–June growing window rather than as an effect specific to a defined phenological stage.

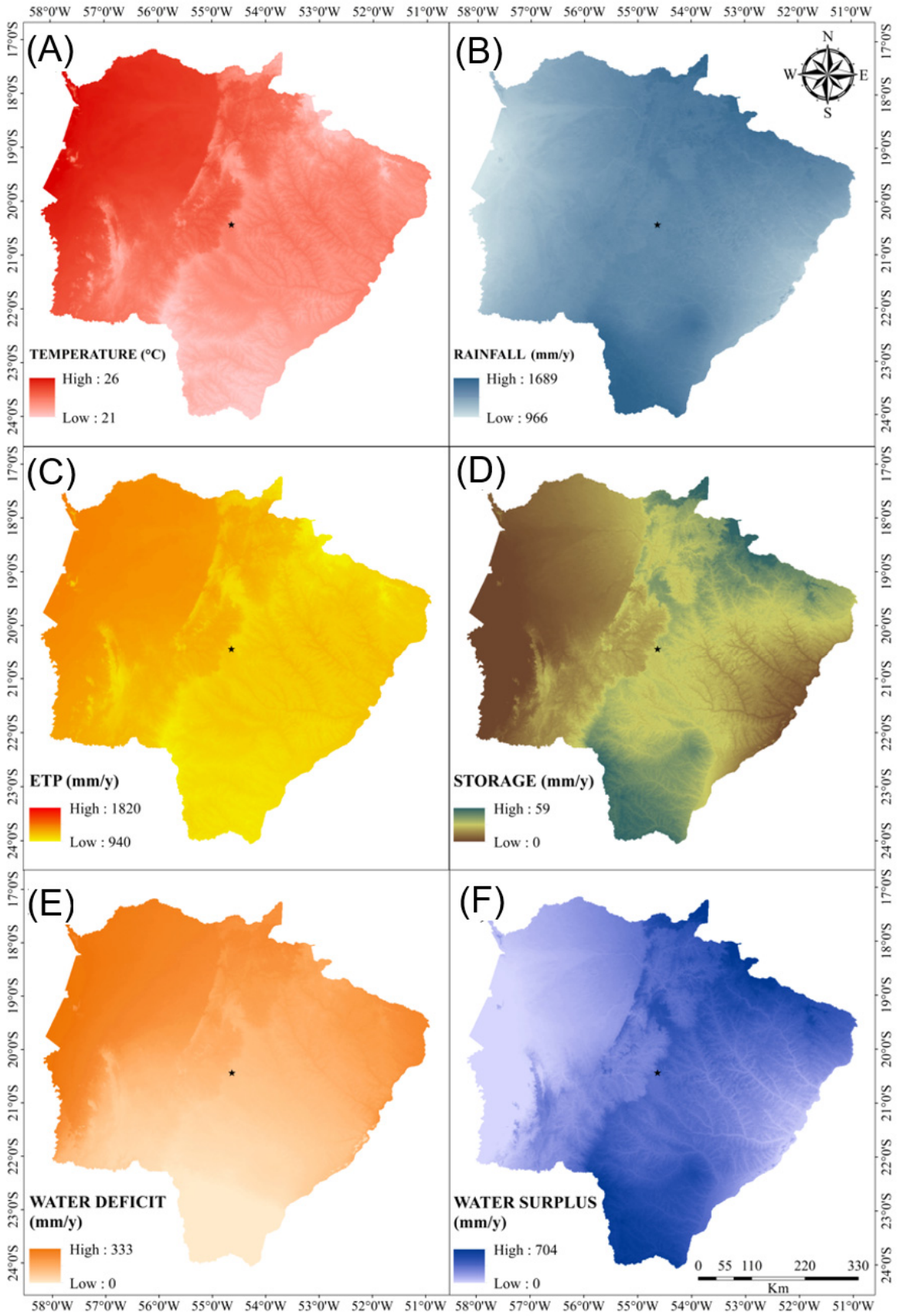


Figure 3. Spatial variability in climate elements during the second-crop corn growing season, based on the average for the period from 2009 to 2023 in the State of Mato Grosso do Sul, Brazil: (**A**) air temperature, (**B**) rainfall, (**C**) evapotranspiration, (**D**) soil water storage, (**E**) water deficit, and (**F**) water surplus.

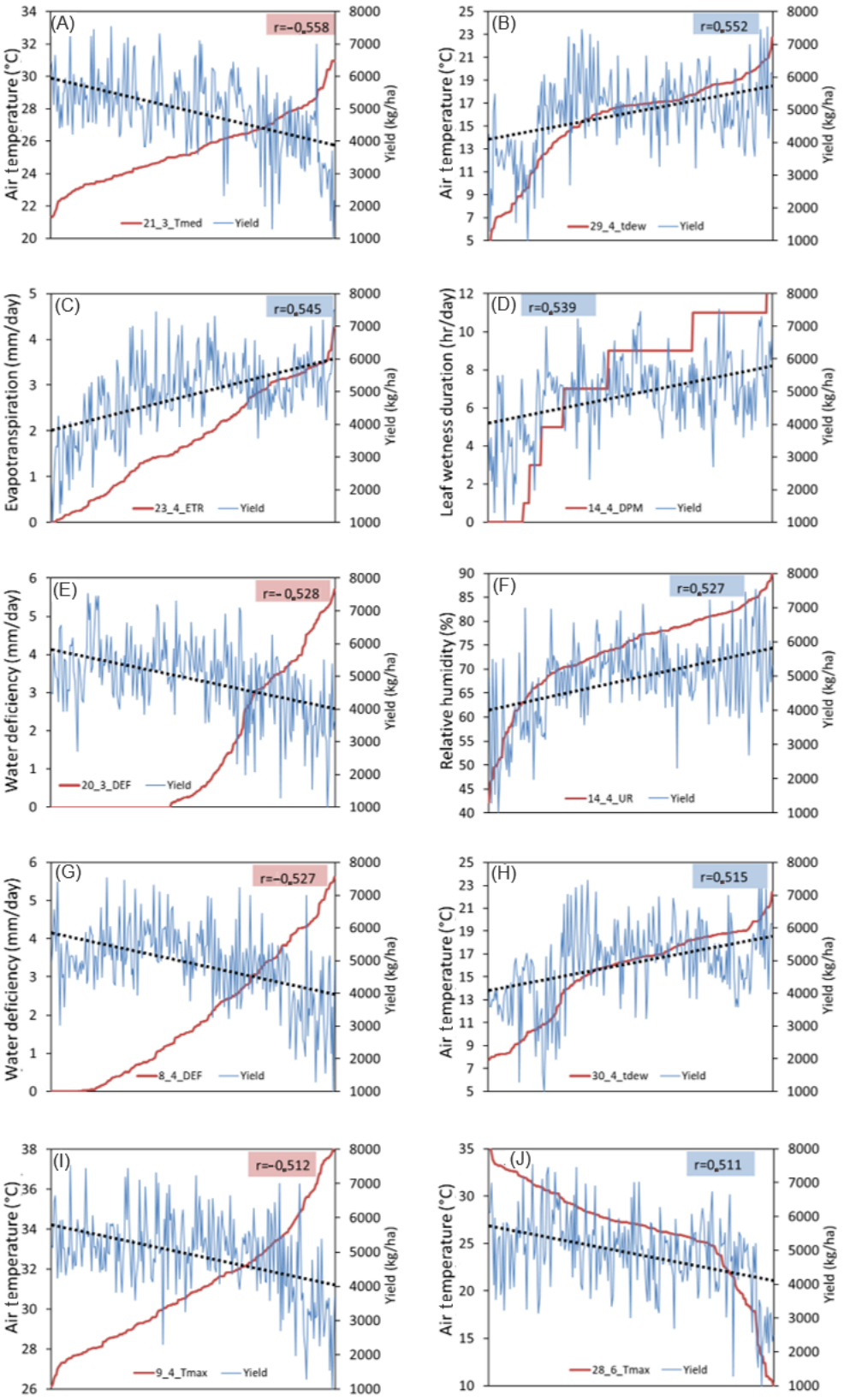


Figure 4. Relationship between climate variables and second-crop corn yield in Mato Grosso do Sul. The x-axis represents the temporal sequence of observations throughout the study period (2012–2023), while the y-axis shows the values of each variable and yield (kg ha^{-1}). The blue lines represent the observed corn yield, the red lines represent a climatic variable showing one of the strongest Pearson correlations with yield (ordered according to its temporal progression), and the black dashed lines indicate the yield trend over time. Panels show: (**A**) mean temperature (March); (**B**) dew temperature

(April); (**C**) actual evapotranspiration (April); (**D**) leaf wetness (April); (**E**) water deficit (March); (**F**) relative humidity (April); (**G**) water deficit (April); (**H**) dew temperature (April); (**I**) maximum temperature (April); and (**J**) maximum temperature (June)..

Furthermore, the correlations of water deficit in March (−0.528) (Figure 4E) and April (−0.527) (Figure 4G) show that water deficit in March and April was negatively associated with second-crop corn yield. These relationships should be interpreted as climatic associations within the March–June growing window rather than as effects specific to a defined phenological stage [63]. Similarly, high maximum temperatures in April were negatively associated with yield within the March–June growing window [64–66].

In contrast, positive correlations (0.552 and 0.515, respectively) associated with dew temperature in April may reflect more favorable atmospheric moisture conditions during the March–June growing window. Because exact phenological dates were not available, these relationships should not be attributed specifically to the R1 stage or to any other defined phenological stage [67,68].

Although advanced explainability techniques such as SHAP values and permutation feature importance were not applied in the present study, Pearson correlation analysis provided a preliminary assessment of associations between climatic variables and second-crop corn yield variability. Variables associated with air temperature, evapotranspiration, relative humidity, and water deficit showed the strongest relationships with yield, indicating their important role in crop development and productivity. Future studies may incorporate explainable artificial intelligence approaches to further improve model interpretability and reduce the "black box" effect commonly associated with machine learning techniques.

Therefore, the present study does not provide model-level attribution of individual feature contributions or directions. The interpretation presented here should be understood as an agronomic and correlation-based assessment of climatic associations with yield, rather than as a direct explanation of the internal decision structure of the machine learning models.

Because some climatic variables are physically interrelated, the Pearson coefficients reported here should be interpreted as marginal associations with yield rather than as independent or causal effects of individual predictors. In particular, relationships involving temperature, evapotranspiration, relative humidity, and water deficit may partially reflect shared climatic information. Therefore, the correlation analysis was used primarily to identify climatic patterns associated with yield variability, while the individual contribution of correlated predictors should be interpreted with caution.

Actual evapotranspiration on April 23 (r = 0.545) was positively associated with yield [69]. Moreover, relative humidity on April 14 (r = 0.527) was also positively associated with yield.

From an agronomic perspective, these relationships are consistent with the climatic sensitivity of second-crop corn grown during the transition from the rainy to the drier season in Mato Grosso do Sul. The negative relationship between air temperature and yield suggests that warmer conditions may intensify crop water demand and accelerate plant development, particularly when high temperatures coincide with limited soil water availability, which may reflect less favorable thermal and water conditions within the March–June growing window. Conversely, the positive relationships of actual evapotranspiration and atmospheric humidity with yield may indicate periods of greater water availability and lower atmospheric water stress, allowing plants to maintain transpiration, photosynthetic activity, and reproductive development. These climatic associations should be interpreted within the broader March–June crop-development window considered in this study.

These findings highlight the complexity of climate interactions in agriculture and the need for agricultural management approaches that consider these variables to mitigate negative impacts and enhance favorable conditions. Understanding these dynamics is vital to optimizing planting planning, selecting more resistant varieties, and implementing efficient water resource management practices, aiming for sustainability and maximizing agricultural yield in Mato Grosso do Sul.

The comparative analysis of the performance of different machine learning models for estimating second-crop corn yield within the 2012–2023 historical data domain in Mato Grosso do Sul (Table 2) illuminates the nuances intrinsic to the different algorithms and their abilities to navigate the complex structure of agricultural data. This exploration reveals not only the methodological variability in the models but also highlights their adaptability and effectiveness in facing the challenges presented by agricultural data, which are notoriously difficult due to their high dimensionality and seasonal variability and the influence of environmental factors. The variation in MAPE and R2 values between models underlines how each one processes and learns from the available data.

Table 2. Calibration, randomly held-out testing performance metrics and optimized hyperparameters of machine learning models for second-crop corn yield prediction in Mato Grosso do Sul, Brazil.

Model	Best Parameters	MAPE (Calibration)	R^2 (Calibration)	MAPE (Test)	R^2 (Test)
MLR	k_features = 14; cv = 2	9.94	0.7281	12.91	0.5952
SVM	kernel = poly; C = 1000; epsilon = 100; gamma = scale; degree = 3	3.38	0.97	14.45	0.76
MLP	hidden_layer_sizes = (10, 11, 12); activation = relu; solver = lbfgs; alpha = 5; learning_rate = constant	5.39	0.96	12.97	0.68
RF	max_depth = 10; n_estimators = 100	4.72	0.97	14.44	0.80
XGBoost	max_depth = 15; n_estimators = 300; learning_rate = 0.2; alpha = 10; colsample_bytree = 0.3	0.15	0.9994	14.05	0.80
GBR	max_depth = 5; n_estimators = 100; random_state = 30	12.81	0.9994	547.78	0.7441

In addition to traditional performance metrics such as MAPE and R^2, model behavior under the adopted random partition was assessed by comparing performance between the calibration and randomly held-out testing subsets. Because this partition did not provide temporal or spatial independence, differences between calibration and testing should be interpreted as indicators of model stability within the present historical dataset, rather than as evidence of robustness to future years or previously unseen locations. Models presenting substantially higher accuracy during calibration than during testing were considered more susceptible to overfitting, indicating reduced capacity to generalize under unseen conditions. This behavior was particularly observed in some highly flexible models, such as MLP and XGBoost, which achieved near-perfect fitting during calibration but showed reduced predictive performance in the randomly held-out testing subset. In contrast, Random Forest demonstrated a smaller calibration-to-testing performance gap than the other highly flexible models, indicating greater stability under the specific data partition used in this study. Because repeated cross-validation, multiple random partitions, and variability measures such as the standard deviation or coefficient of variation were not computed, these results should not be interpreted as a formal statistical demonstration of superior model robustness. This comparative evaluation was therefore used only to characterize

model behavior under the adopted random partition and should not be interpreted as a formal assessment of robustness under independent temporal or spatial conditions.

Multiple linear regression (MLR), with optimized parameters for k_features = 14 and cv = 2, demonstrated an R^2 of 0.7281 in the calibration and 0.5952 in the testing, with MAPE values of 9.94 and 12.91%, respectively (Table 2). These results evidence the model's limitations in addressing the nonlinear complexity of the data since its linear nature restricts its ability to capture more complex relationships between climate variables and corn productivity, which are crucial in variable agricultural contexts [70,71].

The support vector machine (SVM), adjusted with parameters C = 1000, coef0 = 1, degree = 3, epsilon = 100, gamma = scale, and kernel = poly, records an R^2 that varies from 0.97 in the calibration to 0.76 in the randomly held-out testing subset, with a test MAPE of 14.45%. The need to fine-tune parameters and the choice of kernel type directly influence its ability to model data complexity, which may explain the variability in performance and the difficulty in making consistent predictions [72].

The multilayer perceptron (MLP), configured with hidden_layer_sizes = (10, 11, 12), activation = relu, solver = lbfgs, alpha = 5, learning_rate = constant, learning_rate_init = 0.001, and max_iter = 1000, initially stands out with an R2 of 0.96 in the calibration, reducing to 0.68 in the testing, and MAPE values increasing from 5.39 to 12.97%. The drop in performance on testing data suggests overfitting, a common challenge in deep neural networks [73].

Random Forest, with max_depth = 10 and n_estimators = 100, achieved R^2 = 0.97 during calibration and R^2 = 0.80 on the randomly held-out testing subset, with a testing MAPE of 14.44%. Under the adopted random partition, RF showed a smaller calibration-to-testing performance gap than the other highly flexible models. This result indicates greater stability within the present dataset but should not be interpreted as evidence of superior temporal or spatial robustness [35,74,75].

XGBoost achieved the highest calibration accuracy, with a MAPE of 0.15 and R^2 of 0.9994. However, on the randomly held-out testing subset, model performance decreased substantially (MAPE = 14.05 and R^2 = 0.80), indicating evidence of overfitting. Although XGBoost demonstrated strong predictive capability, Random Forest showed more stable behavior between calibration and testing datasets, suggesting greater stability under the randomly held-out observations of the present historical dataset [73,76].

The near-perfect performance of XGBoost during calibration, followed by a substantial reduction in performance on the randomly held-out testing subset, indicates that the model may have fitted specific patterns and noise present in the calibration dataset. This behavior may be related to the high flexibility and complexity of the algorithm, particularly when applied to a dataset with substantial climatic variability and a relatively limited number of observations. Therefore, model performance was interpreted primarily on the basis of the randomly held-out testing subset results, which provide an assessment of predictive performance for held-out observations within the 2012–2023 historical data domain.

Because learning-curve and early-stopping analyses were not performed, the present results do not demonstrate that the high calibration accuracy of XGBoost reflects a generalization advantage; rather, its larger calibration-to-testing gap should be interpreted as evidence of overfitting.

The relatively limited sample size should also be considered when interpreting the performance of the more complex models. The complete dataset comprised 264 municipality-year observations, of which approximately 70% were used for calibration. Under this sample size, highly flexible configurations, such as XGBoost with 300 estimators and the MLP architecture with three hidden layers, may have sufficient capacity to fit idiosyncratic patterns of the calibration data, thereby increasing the risk of overfitting. Learning-curve

analysis and data augmentation were not performed, and simplified model architectures were not evaluated as a separate experiment. Therefore, the present results do not establish that the selected model complexity is optimal for the available sample size. Future studies using larger datasets should compare simpler architectures, regularization strategies, and learning curves to determine whether similar predictive performance can be achieved with lower model complexity.

Although Random Forest and XGBoost achieved similar predictive performance on the randomly held-out testing subset ($R^2 \approx 0.80$), Random Forest showed a smaller calibration-to-testing performance gap. This behavior indicates greater model stability and suggests that Random Forest was less sensitive to the specific characteristics of the calibration dataset, resulting in more consistent performance on the randomly held-out testing subset. Furthermore, GBR, configured with max_depth = 5, n_estimators = 100, and random_state = 30, showed a calibration MAPE of 12.81% and a markedly higher testing MAPE of 547.78%, while the testing R^2 was 0.7441. This large discrepancy indicates that GBR did not provide reliable predictive performance under the present dataset, data partition, and hyperparameter configuration. Because no additional diagnostic experiments were performed to isolate the source of this behavior, the extreme testing MAPE cannot be attributed specifically to overfitting, scaling, implementation, or optimization issues. Therefore, the GBR result is reported transparently but is not used to support the main conclusions regarding model performance. This finding should be interpreted as specific to the present dataset and configuration and not as evidence that the GBR algorithm is intrinsically unsuitable for crop-yield prediction [77–80].

The extremely high testing MAPE obtained by GBR indicates a severe loss of predictive reliability and a failure to generalize to unseen observations, despite its very high calibration performance. This behavior suggests strong sensitivity of the model to the calibration dataset and possible overfitting. Although the available analysis does not allow the specific cause of this instability to be determined, the magnitude of the testing error clearly indicates that GBR was not reliable for yield prediction in this dataset. Therefore, its performance was considered substantially inferior to that of Random Forest and XGBoost under randomly held-out testing subset conditions.

No formal statistical comparison of the calibration and testing distributions was performed. Because the dataset was randomly partitioned at the observation level, both subsets were drawn from the same 2012–2023 historical domain; however, random partitioning may still produce differences in the representation of extreme climatic or yield conditions. Therefore, the exceptionally high testing MAPE of GBR may reflect sensitivity to specific observations, limited robustness to distributional variability, and overfitting to the calibration data. Based on the present results, GBR should not be considered reliable for this dataset and model configuration, although this result should not be interpreted as evidence that the algorithm is intrinsically unsuitable for crop-yield prediction in general.

The comparison between calibration and testing metrics showed that highly flexible models, particularly XGBoost and MLP, presented larger calibration-to-testing performance gaps, indicating a greater risk of overfitting. Random Forest showed a smaller performance gap under the adopted random partition. Accordingly, RF was considered more stable within the present dataset; however, this finding should not be interpreted as formal evidence of superior robustness or generalization to future years or previously unseen locations.

An additional analysis of model stability and generalization ability was performed by comparing the performance metrics obtained during the calibration and testing stages. Models presenting very high accuracy during calibration but substantially lower performance during testing were considered more susceptible to overfitting, indicating limited

capacity to generalize under unseen conditions. This behavior was particularly observed for MLP and XGBoost, which achieved near-perfect fitting during calibration but showed reduced predictive performance in the randomly held-out testing subset. In contrast, Random Forest demonstrated more stable behavior between calibration and testing metrics, indicating greater stability within the specific random partition used in this study. These results highlight the importance of evaluating machine learning models not only on the basis of calibration accuracy but also according to their performance in the randomly held-out testing subset within the 2012–2023 historical data domain.

The map in Figure 5 presents the spatial distribution of observed and predicted corn yield in Mato Grosso do Sul. The observed yield (Figure 5A) corresponds to the average from 2012 to 2023, representing a 12-year data aggregation. Similarly, the predicted yield (Figure 5B) represents the average values estimated by the XGBoost model over the same period. The use of this temporal average aims to reduce the influence of extreme climatic events from specific years, allowing the identification of more stable and representative spatial patterns.

For example, model-estimated yield values were close to the observed yield values in some municipalities, such as Aral Moreira. However, the difference between model-estimated and observed values is higher in other cities, such as Chapadão do Sul and São Gabriel do Oeste, which suggests that the model may not be considering all factors that affect yield in these areas.

The larger discrepancies observed in municipalities such as Chapadão do Sul and São Gabriel do Oeste may reflect the influence of factors not explicitly represented in the model. The yield data used in this study were aggregated at the municipal level, while the predictive models were mainly based on climatic variables. Therefore, local differences in soil characteristics, crop management, cultivar selection, fertilization, pest and disease control, and technological level may have contributed to the observed deviations. In addition, the spatial resolution of the NASA/POWER climatic data may not fully capture local-scale meteorological variability within each municipality. These factors may partly explain why model performance varied spatially, even when the overall predictive accuracy was satisfactory.

An additional source of uncertainty is the spatial-scale mismatch between the response and predictor data. Corn yield was represented at the municipality level, whereas climatic predictors were obtained from a single $0.25^\circ \times 0.25^\circ$ NASA/POWER grid cell associated with each municipality, creating a spatial-scale mismatch. Consequently, within-municipality heterogeneity in climate, soil, topography, and crop management is not explicitly represented. This limitation may be particularly relevant for municipalities with large territorial areas, where a single climatic grid cell may not adequately characterize the environmental conditions of all cultivated areas. The interpolated maps should therefore be interpreted at a regional scale rather than as field- or farm-scale representations.

Accordingly, this mismatch between the spatial scale of the response and predictor data introduces additional uncertainty and limits interpretation of the results to the regional municipal scale.

Because SHAP or other local explainability methods were not applied, it was not possible to attribute the larger prediction errors observed in individual municipalities to specific climatic predictors. Consequently, the spatial discrepancies were interpreted in terms of model limitations and potentially omitted agronomic and environmental factors rather than as feature-specific model effects.

Model performance within the 2012–2023 historical data domain was assessed by comparing model-estimated and observed yield values, but we need to keep in mind that agriculture is affected by several factors, such as cultivation practices and soil conditions.

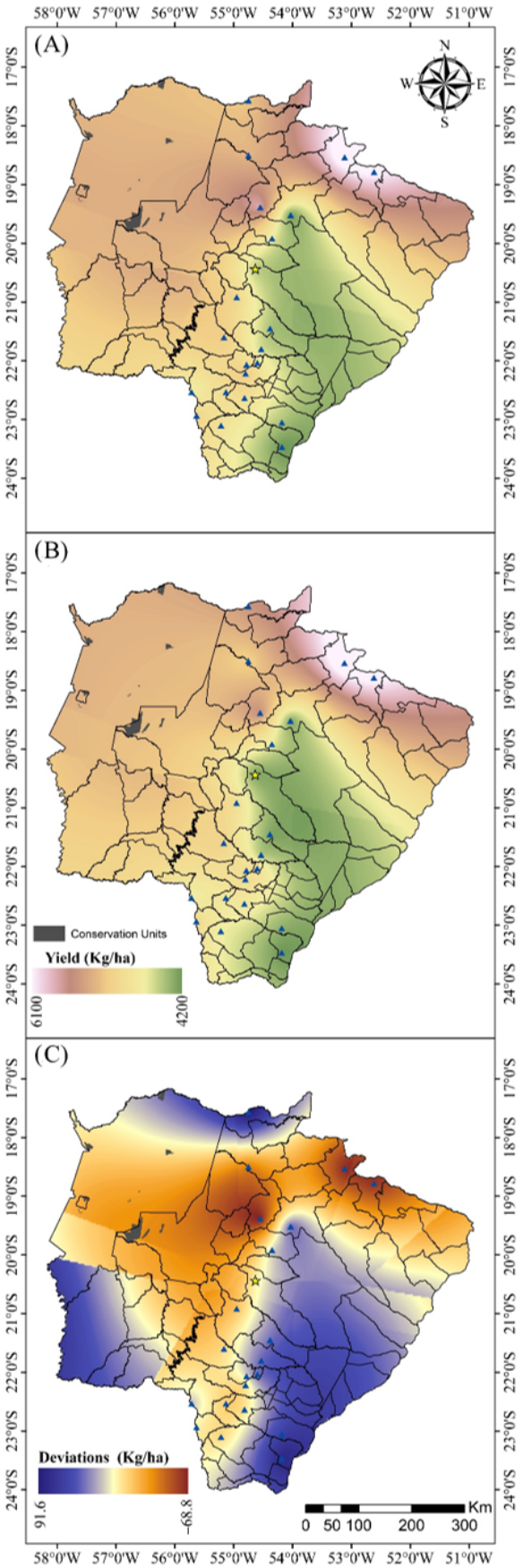


Figure 5. Spatial variation in mean second-crop corn yield in Mato Grosso do Sul for the 2012–2023 period. (**A**) Mean observed municipal yield; (**B**) mean yield predicted by the XGBoost model; (**C**) difference between mean observed and predicted yield. The star indicates the state capital,

while the triangles represent the municipalities with the highest second-crop corn production. The surfaces were generated by Kriging interpolation of municipality-level values, were not subjected to formal geostatistical cross-validation, and should be interpreted only as exploratory regional spatial representations.

An important limitation of the present study is that the prediction models were primarily based on climatic and water-related variables. Other relevant determinants of corn yield, including soil properties, cultivar, planting date, fertilization, pest and disease management, and other agricultural practices, were not explicitly included as predictors.

In addition, the climatological water balance assumed a uniform AWC of 40 mm for all municipalities. In reality, plant-available water capacity varies according to soil texture, structure, effective rooting depth, and other soil hydraulic characteristics. Therefore, the use of a single AWC value may have introduced uncertainty into the estimates of soil water storage, water deficit, and water surplus, particularly in municipalities with contrasting soil conditions. Accordingly, STO, DEF, and EXC should be interpreted as relative indicators derived from the simplified water-balance framework rather than as municipality-specific estimates of actual soil water conditions. Future studies should incorporate spatially explicit soil information and municipality-specific water-holding capacity whenever such data are available.

Although some of these effects may be indirectly reflected in the observed municipal yield data, their individual contributions could not be isolated. This limitation may partly explain the spatial differences between observed and predicted yields and should be considered when interpreting the predictive performance of the models.

Despite some discrepancies, the results obtained with models such as XGBoost illustrate their potential for regional assessment within the analyzed historical dataset. Further temporally and spatially independent validation would be required before considering operational forecasting applications.

Figure 5C shows a map of the deviations between the observed and model-estimated XGBoost corn yield values. Importantly, Camapuã had a positive deviation of 290 kg ha^{-1}, indicating that the actual yield was much higher than expected. This may be due to an exceptionally good agricultural year, the use of more advanced cultivation technologies, or an initial underestimation of potential yield. On the other hand, São Gabriel do Oeste had a negative deviation of -328 kg ha^{-1}.

Although correlation-based screening was used to reduce redundancy among climatic predictors, some residual multicollinearity may remain, particularly among temperature, evapotranspiration, humidity, and water-deficit variables, and should be further investigated in future studies.

4. Conclusions

Second-crop corn yield in Mato Grosso do Sul was associated with climatic and water-related variability during the March–June growing window. Higher temperatures and water deficits were generally associated with lower yield, whereas greater atmospheric humidity and actual evapotranspiration were associated with more favorable yield conditions. These results should be interpreted as climatic associations over the analyzed growing period and not as evidence of effects specific to particular phenological stages.

The spatial and water-balance results should also be interpreted with caution. The uniform AWC value of 40 mm was a simplified assumption that was not adjusted for municipality-specific soil properties, and therefore STO, DEF, and EXC should be regarded as relative indicators. In addition, municipal yield values were compared with climatic information derived from a single $0.25^\circ \times 0.25^\circ$ grid cell per municipality, resulting in a

spatial-scale mismatch. The Kriging maps were not subjected to formal geostatistical cross-validation and should therefore be interpreted only as exploratory spatial representations.

Among the evaluated machine learning approaches, Random Forest and XGBoost showed the strongest predictive performance on the randomly held-out testing subset, both reaching R^2 values of approximately 0.80. However, XGBoost exhibited near-perfect calibration performance followed by a substantial decrease during testing, indicating overfitting. Random Forest showed a smaller calibration-to-testing performance gap and was therefore considered the preferred model under the present random partition. This conclusion is strictly limited to the current dataset and partition. Because temporal and spatial independence were not imposed during validation, the reported performance metrics should not be interpreted as evidence of predictive performance for unseen future years or new geographic regions.

Temporal or spatial extrapolation would require validation strategies such as Leave-One-Year-Out, temporally ordered validation, or spatial cross-validation, which were not performed in the present study.

The exploratory spatial representations showed broad regional agreement between observed and model-estimated corn yield values, although differences between observed and predicted values in some municipalities indicate that climatic variables alone cannot explain all local yield variability. Factors such as soil characteristics, cultivar, sowing date, fertilization, pest and disease management, and other agricultural practices may also contribute to these differences. Therefore, the modeling results provide a regional assessment of areas associated with climatic limitations within the analyzed historical dataset. Future studies incorporating additional production-related variables and explainable artificial intelligence techniques may further improve prediction accuracy and model interpretation.

Overall, the findings should be interpreted as a regional assessment of established machine learning methods within the 2012–2023 historical data domain and under the stated data and validation limitations. The study does not introduce a new machine learning algorithm. The reported predictions refer to estimates for held-out observations within the study dataset and the adopted random partition, rather than to operational or future-year forecasting.

Author Contributions: Conceptualization, G.B.T., F.G.M. and L.E.O.A.; methodology, G.B.T.; validation, G.B.T. and L.E.O.A.; formal analysis, G.B.T.; investigation, G.B.T., F.G.M. and L.E.O.A.; resources, G.B.T.; data curation, G.B.T.; writing—original draft preparation, G.B.T., F.G.M. and L.E.O.A.; writing—review and editing, G.B.T., F.G.M. and L.E.O.A.; visualization, L.E.O.A.; funding acquisition, G.B.T. All authors have read and agreed to the published version of the manuscript.

Funding: This study received funding from FUNDECT (Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia of the State of Mato Grosso do Sul, Grant Term No. 265/2022; SIAFEM: 32183).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data will be made available upon request.

Acknowledgments: We extend our thanks to the IFMS (Federal Institute of Mato Grosso do Sul) for its collaboration and support throughout the process.

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

References

1. Fussell, B.H. *The Story of Corn*; University of New Mexico Press: Albuquerque, NM, USA, 2004; Available online: https://www.unmpress.com/9780826335920/the-story-of-corn/ (accessed on 21 March 2025).
2. Mangelsdorf, P.C. *Corn: Its Origin, Evolution and Improvement*; Harvard University Press: Cambridge, MA, USA, 1974.
3. Erenstein, O.; Jaleta, M.; Sonder, K.; Mottaleb, K.; Prasanna, B.M. Global maize production, consumption and trade: Trends and R&D implications. *Food Secur.* **2022**, *14*, 1295–1319. [CrossRef]
4. United States Department of Agriculture (USDA); Foreign Agricultural Service (FAS). *Brazil: Grain and Feed Update (Report BR2026-0031)*; USDA: Washington, DC, USA, 2026. Available online: https://apps.fas.usda.gov/newgainapi/api/Report/DownloadReportByFileName?fileName=Grain+and+Feed+Update_Brasilia_Brazil_BR2026-0031.pdf (accessed on 25 August 2026).
5. Instituto Brasileiro de Geografia e Estatística (IBGE). *Levantamento Sistemático da Produção Agrícola*; IBGE: Rio de Janeiro, Brazil, 2022. Available online: https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9201-levantamento-sistematico-da-producao-agricola.html (accessed on 1 October 2023).
6. Landau, E.C.; Hirsch, A.; Cruz, J.C.; Duarte, J.O.; Garcia, J.C.; Garagorry, F.L. *Áreas de Concentração da Produção de Milho no Brasil entre 2008 e 2010*; Embrapa Milho e Sorgo: Sete Lagoas, MG, Brazil, 2012. Available online: https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/950094/1/bol58.pdf (accessed on 1 October 2023).
7. United States Department of Agriculture (USDA); Foreign Agricultural Service (FAS). *Grain and Feed Update: Brazil (Report BR2023-0028)*; USDA: Brasília, Brazil, 2023. Available online: https://apps.fas.usda.gov/newgainapi/api/Report/DownloadReportByFileName?fileName=Grain+and+Feed+Update_Brasilia_Brazil_BR2023-0028.pdf (accessed on 5 October 2024).
8. Oliveira, A.L.R.; Silveira, J.M. Restructuring of the corn supply chain in Brazil: Facing the challenges in logistics or regulation of biotechnology. *Int. Food Agribus. Manag. Rev.* **2013**, *16*, 1–24. [CrossRef]
9. De Faria, C.H.F.; Almeida, J.F.F.; Pinto, L.R. Simulation–optimisation approach for sustainable planning of intermodal logistics in the Brazilian grain export industry. *Decis. Anal. J.* **2024**, *10*, 100388. [CrossRef]
10. Ferreira, J.V. Geração e Distribuição de Riqueza no Âmbito da Cultura do Milho Primeira Safra nos Principais Estados Produtores do Brasil. 2023. Available online: https://repositorio.ufu.br/bitstream/123456789/36896/1/Gera%C3%A7%C3%A3oDistribui%C3%A7%C3%A3oRiqueza.pdf (accessed on 5 October 2024).
11. Martinelli, L.A.; Naylor, R.; Vitousek, P.M.; Moutinho, P. Agriculture in Brazil: Impacts, costs, and opportunities for a sustainable future. *Curr. Opin. Environ. Sustain.* **2010**, *2*, 431–438. [CrossRef]
12. Sexton, S.E.; Martin, L.A.; Zilberman, D. Biofuel and biotech: A sustainable energy solution? *ARE Update* **2006**, *9*, 1–4.
13. Companhia Nacional de Abastecimento (CONAB). *Acompanhamento da Safra Brasileira de Grãos 2022/23*; CONAB: Brasília, Brazil, 2023. Available online: https://www.gov.br/conab/pt-br/atuacao/informacoes-agropecuarias/safras/safra-de-graos (accessed on 5 October 2024).
14. Lobell, D.B.; Gourdji, S.M. The influence of climate change on global crop productivity. *Plant Physiol.* **2012**, *160*, 1686–1697. [CrossRef] [PubMed]
15. Hatfield, J.L.; Dold, C. Climate change impacts on corn phenology and productivity. In *Corn: Production and Human Health in Changing Climate*; IntechOpen: London, UK, 2018; pp. 95–114.
16. Sánchez, B.; Rasmussen, A.; Porter, J.R. Temperatures and the growth and development of maize and rice: A review. *Glob. Change Biol.* **2014**, *20*, 408–417. [CrossRef] [PubMed]
17. Galon, L.; Tironi, S.P.; Rocha, A.A.; Soares, E.R.; Concenço, G.; Alberto, C.M. Influência dos fatores abióticos na produtividade da cultura do milho. *Rev. Trópica Ciênc. Agrár. Biol.* **2010**, *4*, 18–38.
18. Hoogenboom, G. Contribution of agrometeorology to the simulation of crop production and its applications. *Agric. For. Meteorol.* **2000**, *103*, 137–157. [CrossRef]
19. Bojanowski, J.S.; Sikora, S.; Musiał, J.P.; Woźniak, E.; Dąbrowska-Zielińska, K.; Slesiński, P.; Milewski, T.; Łączyński, A. Integration of Sentinel-3 and MODIS vegetation indices with ERA-5 agro-meteorological indicators for operational crop yield forecasting. *Remote Sens.* **2022**, *14*, 1238. [CrossRef]
20. Zhuo, W.; Fang, S.; Gao, X.; Wang, L.; Wu, D.; Fu, S.; Wu, Q.; Huang, J. Crop yield prediction using MODIS LAI, TIGGE weather forecasts and WOFOST model: A case study for winter wheat in Hebei, China during 2009–2013. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *106*, 102668. [CrossRef]
21. Bowen, W.T.; Thornton, P.K.; Hoogenboom, G. The simulation of cropping sequences using DSSAT. In *Understanding Options for Agricultural Production*; Tsuji, G.Y., Hoogenboom, G., Thornton, P.K., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1998; pp. 313–327. [CrossRef]
22. Lekakis, E.; Zaikos, A.; Polychronidis, A.; Efthimiou, C.; Pourikas, I.; Mamouka, T. Evaluation of different modelling techniques with fusion of satellite, soil and agro-meteorological data for the assessment of durum wheat yield. *Agriculture* **2022**, *12*, 1635. [CrossRef]

23. Debalke, D.B.; Abebe, J.T. Maize yield forecast using GIS and remote sensing in Kaffa Zone, South West Ethiopia. *Environ. Syst. Res.* **2022**, *11*, 1. [CrossRef]
24. Ray, M.; Singh, K.N.; Pal, S.; Saha, A.; Sinha, K.; Kumar, R.R. Rainfall prediction using time-delay wavelet neural network (TDWNN) model for assessing agrometeorological risk. *J. Agrometeorol.* **2023**, *25*, 151–157. [CrossRef]
25. Alves de Lima, M.J.; Nunes, H.G.G.C.; Sampaio, L.S.; Souza, P.J.d.O.P.d.; Fraisse, C.W. Optimal soybean sowing window adjusted to climatic variability based on El Niño-Southern Oscillation using agrometeorological modeling. *Pesqui. Agropecu. Trop.* **2022**, *52*, e72428. [CrossRef]
26. Faria Junior, C.A.; de Freitas, P.S.L.; Dallacort, R.; Mioto, L.S.; Barbieri, J.D. Agricultural zoning of popcorn corn for the state of Mato Grosso, Brazil. *Appl. Res. Agrotechnol.* **2020**, *13*, 1–13.
27. Thornthwaite, C.W.; Mather, J.R. *The Water Balance*; Drexel Institute of Technology: Centerton, NJ, USA, 1955. Available online: https://books.google.com/books?id=DTdtcgAACAAJ (accessed on 1 October 2023).
28. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration—Guidelines for Computing Crop Water Requirements*; FAO Irrigation and Drainage Paper 56; FAO: Rome, Italy, 1998. Available online: https://www.fao.org/4/x0490e/x0490e00.htm (accessed on 2 October 2023).
29. Faramiñán, M.A.; Laino, C.; Carmona, F.; Holzman, M.E.; Rivas, R.E. Evaluation of a generalized linear model for actual evapotranspiration using satellite and reanalysis data. In *Proceedings of the IEEE Biennial Congress of Argentina (ARGENCON), San Juan, Argentina, 7–9 September 2022*; IEEE: Piscataway, NJ, USA, 2022; pp. 1–5.
30. Huntingford, C.; Jeffers, E.S.; Bonsall, M.B.; Christensen, H.M.; Lees, T.; Yang, H. Machine learning and artificial intelligence to aid climate change research and preparedness. *Environ. Res. Lett.* **2019**, *14*, 124007. [CrossRef]
31. Aparecido, L.E.O.; Padua, J.M.V.; Torsoni, G.B.; Barboza, T.O.C.; Viol, L.E.S.; Moraes, J.R.d.S.C.d.; dos Santos, A.F. Climate risk zoning for wheat crops in the southeastern region of Brazil. *J. Sci. Food Agric.* **2024**, *104*, 456–467. [CrossRef] [PubMed]
32. De Sá Júnior, A.; de Carvalho, L.G.; da Silva, F.F.; Alves, M.d.C. Application of the Köppen classification for climatic zoning in Minas Gerais, Brazil. *Theor. Appl. Climatol.* **2012**, *108*, 1–7. [CrossRef]
33. Ilyas, Q.M.; Ahmad, M.; Mehmood, A. Automated estimation of crop yield using artificial intelligence and remote sensing technologies. *Bioengineering* **2023**, *10*, 125. [CrossRef] [PubMed]
34. de Oliveira Aparecido, L.E.; Lorençone, P.A.; Lorençone, J.A.; de Meneses, K.C.; de Lima, R.F.; da Silva Cabral Moraes, J.R.; Torsoni, G.B. Coffee pest severity by agrometeorological models in subtropical climate. *Int. J. Biometeorol.* **2022**, *66*, 957–969. [CrossRef] [PubMed]
35. Smith, M.J. Getting value from artificial intelligence in agriculture. *Anim. Prod. Sci.* **2024**, *60*, 46–54.
36. Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson correlation coefficient. In *Noise Reduction in Speech Processing*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–4. [CrossRef]
37. Morelli-Ferreira, F.; Maia, N.; Tedesco, D.; Kazama, E.; Carneiro, F.; Santos, L.; Seben Junior, G.; Rolim, G.; Shiratsuchi, L.; Silva, R. Comparison of machine learning techniques in cotton yield prediction using satellite remote sensing. *Preprints* **2021**. [CrossRef]
38. Torsoni, G.B.; Aparecido, L.E.d.O.; dos Santos, G.M.; Chiquitto, A.G.; Moraes, J.R.d.S.C.; Rolim, G.d.S. Soybean yield prediction by machine learning and climate. *Theor. Appl. Climatol.* **2023**, *151*, 1709–1725. [CrossRef]
39. Andrea, M.C.S.; Boote, K.J.; Sentelhas, P.C.; Romanelli, T.L. Variability and limitations of maize production in Brazil: Potential yield, water-limited yield and yield gaps. *Agric. Syst.* **2018**, *165*, 264–273. [CrossRef]
40. Zorić, M.; Gunjača, J.; Galić, V.; Jukić, G.; Varnica, I.; Šimić, D. Best linear unbiased predictions of environmental effects on grain yield in maize variety trials. *Agronomy* **2022**, *12*, 922. [CrossRef]
41. Shams, M.Y.; Gamel, S.A.; Talaat, F.M. Enhancing crop recommendation systems with explainable artificial intelligence: A study on agricultural decision-making. *Neural Comput. Appl.* **2024**, *36*, 5695–5714. [CrossRef]
42. Masood, A.; Niazkar, M.; Zakwan, M.; Piraei, R. A machine learning-based framework for water quality index estimation in the Southern Bug River. *Water* **2023**, *15*, 3543. [CrossRef]
43. Jin, Z.; Zhuang, Q.; Tan, Z.; Dukes, J.S.; Zheng, B.; Melillo, J.M. Do maize models capture the impacts of heat and drought stresses on yield? Using algorithm ensembles to identify successful approaches. *Glob. Chang. Biol.* **2016**, *22*, 3112–3126. [CrossRef] [PubMed]
44. Martinez, C.J.; Baigorria, G.A.; Jones, J.W. Use of climate indices to predict corn yields in southeast USA. *Int. J. Climatol.* **2009**, *29*, 1680–1691. [CrossRef]
45. Adisa, O.M.; Botai, C.M.; Botai, J.O.; Hassen, A.; Darkey, D.; Tesfamariam, E.; Adisa, A.F.; Adeola, A.M.; Ncongwane, K.P. Analysis of agro-climatic parameters and their influence on maize production in South Africa. *Theor. Appl. Climatol.* **2018**, *134*, 991–1004. [CrossRef]
46. Butts-Wilmsmeyer, C.J.; Seebauer, J.R.; Singleton, L.; Below, F.E. Weather during key growth stages explains grain quality and yield of maize. *Agronomy* **2019**, *9*, 16. [CrossRef]
47. Schmidt, J.L.; Hallauer, A.R. Estimating harvest date of corn in the field. *Crop Sci.* **1966**, *6*, 227–231. [CrossRef]

48. Kabela, E.D.; Hornbuckle, B.K.; Cosh, M.H.; Anderson, M.C.; Gleason, M.L. Dew frequency, duration, amount, and distribution in corn and soybean. *Agric. For. Meteorol.* **2009**, *149*, 11–24. [CrossRef]
49. Ko, J.; Piccinni, G. Corn yield responses under crop evapotranspiration-based irrigation management. *Agric. Water Manag.* **2009**, *96*, 799–808. [CrossRef]
50. Lorençone, P.A.; de Oliveira Aparecido, L.E.; Lorençone, J.A.; Torsoni, G.B.; de Lima, R.F. Estimation of Air Temperature Using Climate Factors in Brazilian Sugarcane Regions. *Rev. Bras. Meteorol.* **2022**, *37*, 121–140. [CrossRef]
51. Zhou, Y.; Pan, M.; Guan, W.; Fu, C.; Su, T. Predicting sugarcane yield via improved least squares support vector machine. *Agriculture* **2023**, *13*, 2115. [CrossRef]
52. Ma, Y.; Zhang, Z.; Kang, Y.; Özdoğan, M. Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. *Remote Sens. Environ.* **2021**, *259*, 112408. [CrossRef]
53. Bebie, M.; Cavalaris, C.; Kyparissis, A. Assessing durum wheat yield through Sentinel-2 imagery: A machine learning approach. *Remote Sens.* **2022**, *14*, 3880. [CrossRef]
54. Wei, M.C.F.; Maldaner, L.F.; Ottoni, P.M.N.; Molin, J.P. Carrot yield mapping: A precision agriculture approach based on machine learning. *AI* **2020**, *1*, 229–241. [CrossRef]
55. Esparza-Gómez, J.M.; Luque-Vega, L.F.; Guerrero-Osuna, H.A.; Carrasco-Navarro, R.; García-Vázquez, F.; Mata-Romero, M.E.; Olvera-Olvera, C.A.; Carlos-Mancilla, M.A.; Solís-Sánchez, L.O. Long Short-Term Memory Recurrent Neural Network and Extreme Gradient Boosting Algorithms Applied in a Greenhouse's Internal Temperature Prediction. *Appl. Sci.* **2023**, *13*, 12341. [CrossRef]
56. Manjunath, M.C.; Palayyan, B.P. An efficient crop yield prediction framework using hybrid machine learning model. *Rev. Intell. Artif.* **2023**, *37*, 1057–1067. [CrossRef]
57. Sabitov, B.; Kartanova, A.; Uulu, T.K.; Seitkazieva, N.; Dyikanova, A.; Orozobekova, A. Modeling and forecasting tasks of agriculture based on machine learning. *E3S Web Conf.* **2023**, *380*, 01001. [CrossRef]
58. Jhajharia, K.; Mathur, P. Machine learning based crop yield prediction model in Rajasthan region of India. *Iraqi J. Sci.* **2024**, *65*, 390–400. [CrossRef]
59. Naumann, G.; Podesta, G.; Marengo, J.; Luterbacher, J.; Bavera, D.; Arias-Muñoz, C.; Marinho Ferreira Barbosa, P.; Cammalleri, C.; Chamorro, L.; Cuartas, L.A.; et al. *The 2019–2021 Extreme Drought Episode in La Plata Basin*; Publications Office of the European Union: Luxembourg, 2021. [CrossRef] [PubMed]
60. Rao, V.B.; Franchito, S.H.; Rosa, M.B.; Govardhan, D.; Figueroa, S.N.; Bhargavi, V.L. In a changing climate Hadley cell induces a record flood in Amazon and drought in South Brazil in 2021. *Nat. Hazards* **2022**, *114*, 1549–1561. [CrossRef]
61. Dos Reis, C.J.; Souza, A.; Graf, R.; Kossowski, T.M.; Abreu, M.C.; de Oliveira-Junior, J.F.; Fernandes, W.A. Modeling air temperature using extreme value theory in Mato Grosso do Sul. *Stoch. Environ. Res. Risk Assess.* **2022**, *36*, 3499–3516. [CrossRef] [PubMed]
62. Teodoro, P.E.; de Oliveira-Júnior, J.F.; Da Cunha, E.R.; Correa, C.C.G.; Torres, F.E.; Bacani, V.M.; Gois, G.; Pereira Ribeiro, L. Cluster analysis applied to the spatial and temporal variability of monthly rainfall in Mato Grosso do Sul State, Brazil. *Meteorol. Atmos. Phys.* **2016**, *128*, 197–209. [CrossRef]
63. Yang, B.; Wu, S.; Yan, Z. Effects of climate change on corn yields: Spatiotemporal evidence. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 433. [CrossRef]
64. Devastthale, A.; Carlund, T.; Karlsson, K.G. Trends in agrometeorological climate variables. *Agric. For. Meteorol.* **2022**, *316*, 108849. [CrossRef]
65. Dwamena, H.A.; Tawiah, K.; Akuoko, A.S. Effect of rainfall, temperature and relative humidity on the yield of cassava, yam and maize in the Ashanti Region of Ghana. *Int. J. Agron.* **2022**, *2022*, 9077383. [CrossRef]
66. Hatfield, J.L. Increased temperatures effects on maize yield. *Agric. Environ. Lett.* **2016**, *1*, 150006. [CrossRef]
67. Ferreira Nascimento, W.; da Costa, J.S.; Peixoto, P.P.P.; Duarte, N.D.L. Effects of temperature on soybean and corn in Mato Grosso do Sul. *Investig. Agrar.* **2018**, *20*, 30–37. [CrossRef]
68. Zuffo, A.M.; de Oliveira, A.M.; Aguilera, J.G.; Ratke, R.F.; Steiner, F.; de Abreu, C.M.; Fonseca, W.L.; dos Santos, A.S.; Martínez, L.A.; Aranibar, L.M.; et al. Correlations and path analysis of second-crop corn hybrids. *Aust. J. Crop Sci.* **2023**, *17*, 639–644. [CrossRef]
69. Oliveira, E.F. Fatores que garantem o sucesso da safrinha de milho. In *Seminário Nacional de Milho Safrinha*; IAPAR: Londrina, Brazil, 2003; Available online: https://e3sys.com.br/milho/ (accessed on 4 May 2024).
70. Porto, C.B. O seguro agrícola privado no Brasil. *Rev. Política Agríc.* **1999**, *8*, 1–10.
71. Coelho, A.M.; Parentoni, S.N. *Milho Verde*; Embrapa: Brasília, Brazil, 1988.
72. Casela, C.R.; Ferreira, A.S. *A Cercosporiose Na Cultura Do Milho*; Embrapa: Brasília, Brazil, 2003; Available online: https://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/487543 (accessed on 4 May 2024).
73. Mas, D.M.L.; Ahlfeld, D.P. Comparing neural networks and regression models. *Hydrol. Sci. J.* **2007**, *52*, 713–731. [CrossRef]

74. Oliveira Aparecido, L.E.; da Silva Cabral de Moraes, J.R.; de Meneses, K.C.; Torsoni, G.B.; de Lima, R.F.; Costa, C.T.S. Köppen-Geiger and Camargo climate classifications for the Midwest of Brasil. *Theor. Appl. Climatol.* **2020**, *142*, 1133–1145. [CrossRef]
75. De Souza, P.V.D.; de Rezende, L.P.; Duarte, A.P.; Miranda, G.V. Maize yield prediction using artificial neural networks. *Eng. Technol. Appl. Sci. Res.* **2023**, *13*, 10338–10346. [CrossRef]
76. Ahangar, R.G.; Yahyazadehfar, M.; Pournaghshband, H. The comparison of methods artificial neural network with linear regression using specific variables for prediction stock price in Tehran stock exchange. *arXiv* **2010**, arXiv:1003.1457. [CrossRef]
77. Krige, D.G. A statistical approach to some basic mine valuation problems on the Witwatersrand. *J. S. Afr. Inst. Min. Metall.* **1951**, *52*, 119–139.
78. Schaafsma, W.; Van Vark, G.N. Classification and discrimination problems. *Stat. Neerl.* **1979**, *33*, 91–126. [CrossRef]
79. Olisah, C.C.; Smith, L.; Smith, M.; Lawrence, M.O.; Ojukwu, O. Corn yield prediction model with deep neural networks. *arXiv* **2024**, arXiv:2401.03768.
80. Zou, Y.; Kattel, G.R.; Miao, L. Enhancing maize yield simulations using machine learning. *Remote Sens.* **2024**, *16*, 701. [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.