Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5, 1213-1234
2024
Publisher: Learning Gate
DOI: 10.55214/25768484.v8i5.1826

Assessing the impact of climate variability on wheat yield in Bloemfontein wheat farms through time series analysis

Vincent Zibi Mohale[1], Ibidun Christiana Obagbuwa[2*]
[1,2]Department of Computer Science and Information Technology, Faculty of Natural and Applied Sciences, Sol Plaatje University, Kimberley, South Africa; 201902877@spu.ac.za (V.Z.M.) Ibidun.obagbuwa@spu.ac.za (I.C.O.)

Abstract: Climate change, characterized by long-term shifts in global or regional weather patterns, is a consequence of natural processes and human activities. These shifts encompass al-terations in temperature, precipitation, wind patterns, and other climatic variables, all of which exert direct influence on crop growth, development, and overall agricultural productivity. Comprehending the intricate relationship between climate change and crop production is par-amount for formulating strategies to counteract adverse effects and adapt to evolving conditions. This study focuses on assessing the impact of climate variability on wheat yield in Bloemfontein wheat farms through rigorous time series analysis. The research involved the application of various time series models, including SARIMA, ARIMA, Facebook Prophet, LSTM, VAR, and Multiple Linear Regression. The investigation began with forecasting temperature patterns using SARIMA and Facebook Prophet models. SARIMA outperformed Facebook Prophet in this context, as evidenced by lower RMSE and MSE metrics. Subsequently, the study delved into predicting rainfall and precipitation, employing ARIMA and LSTM models. In this case, LSTM demonstrated superior predictive capabilities. Finally, wheat production yield was analyzed using VAR and Multiple Linear Regression, with VAR yielding more accurate predictions. The findings of this study hold profound implications for policymakers, farmers, and stakeholders deeply invested in agriculture and food security. By shedding light on the repercussions of climate change on crop production through the application of time series analysis, this project aspires to contribute to developing sustainable agricultural practices, robust farming systems and proactive policies de-signed to mitigate the adverse effects of climate change on global food production.
Keywords: *Climate; Time series analysis; models; wheat yield, wheat farm, crop production.*

1. Introduction

Climate change poses a formidable challenge to global agricultural systems, with discernible impacts on crop production and food security. In the context of shifting climatic patterns, investigating the effects of climate variability on agricultural yields has garnered substantial scholarly attention [1, 2, 3]. A comprehensive study was conducted to analyse the intricate relationships between temperature anomalies and crop yield fluctuations [1]. Their findings underscored the vulnerability of crops to abnormal temperature spikes, emphasizing the need for localized investigations to grasp the region-specific impacts of climate variability. In parallel, [4] explored the correlation between precipitation patterns and crop productivity, revealing the susceptibility of rain-fed agriculture to irregular rainfall regimes. Extending this discourse, the current study delves into the repercussions of climate variability on wheat yield in Bloemfontein wheat farms. Like many others, the Bloemfontein region confronts the challenge of altering climate conditions that can disrupt agricultural practices and ultimately influence crop production. Understanding the nexus between climate variations and wheat yield within this context is pivotal for devising targeted adaptation strategies. [3] contributed to this body of research

* Correspondence: ibidun.obagbuwa@spu.ac.za

by examining the potential interaction between climate factors and crop yield stability. Using advanced modeling techniques, they showcased the intricate ways temperature, precipitation, and variations impact yield fluctuations, offering insights into predicting future yield trends. Building on such advancements, this study employs a similar methodology to analyze the historical climate data of Bloemfontein and its implications for wheat farming. Amidst growing concerns about food security and the sustainability of agricultural systems, this study fills a crucial research gap by shedding light on how climate variability influences wheat yield in a localized agricultural setting. The insights gained from this investigation hold implications for farmers, policymakers, and stakeholders striving to enhance the resilience of agricultural systems in the face of changing climatic conditions. [4], advanced the field by investigating the direct impacts of climate variability on crop yields and the secondary effects on soil quality and nutrient availability. Their study highlighted the intricate interactions between climate, soil health, and agricultural productivity. This holistic perspective underscores the need for integrated approaches considering biophysical and ecological factors when assessing climate-yield relationships.

The present study aims to bridge the gap between climate science and agricultural practices by focusing on Bloemfontein's wheat farms. The unique characteristics of the Bloemfontein region, including its semi-arid climate and reliance on rain-fed agriculture, make it an ideal case study for understanding the localized consequences of climate variability on wheat yield [1]. By amalgamating insights from [5] and similar research, this study takes a multi-dimensional approach that considers climate variables and their cascading impacts on the overall agricultural ecosystem. [6] enriched the discourse by examining potential mitigation strategies to alleviate the adverse impacts of climate variability on crop yield. They explored the efficacy of precision agriculture techniques and modified cropping calendars to adapt to changing climate conditions. Such proactive strategies could prove essential for Bloemfontein's wheat farmers, who must grapple with uncertain precipitation patterns and temperature fluctuations. To gain a comprehensive understanding of the impact of climate change on crop production, this study aims to employ machine learning algorithms as powerful tools for analysis. The intricate relationship between climatic factors and agricultural productivity will be explored by leveraging historical climate data and crop production records. Machine learning techniques allow us to identify the most influential climate variables, develop predictive models, and assess the potential impacts of climate change on crop yields. By integrating insights from previous studies, such as those by [7] and [8], [9], and [10], the aim is to build upon existing knowledge and provide a robust assessment of the effects of climate change on crop production. The analysis will contribute to the ongoing efforts in formulating strategies to mitigate the negative impacts of climate change and ensure global food security

This investigation contributes to the expanding body of knowledge elucidating the intricate interplay between climate variability and agricultural productivity. By grounding our study in the specific context of Bloemfontein's wheat farms and drawing on the insights of pioneering researchers such as [3] and [6], we aim to provide a nuanced understanding of the challenges and opportunities presented by climate variability. Ultimately, these findings can empower stakeholders to make informed decisions that enhance the resilience and sustainability of wheat farming in Bloemfontein.

2. Materials and Methods

2.1. Time Series Analysis for Agricultural Data

Time series analysis serves as a powerful tool for investigating the temporal dynamics of agricultural data, enabling the identification of patterns, trends, and underlying relationships. In the context of our study on climate variability's impact on wheat yield in Bloemfontein wheat farms, time series analysis provides a structured framework to uncover insights from historical weather and crop yield data. Time series data consists of observations collected at successive time points, creating a sequential data sequence. Applying time series analysis to agricultural data involves exploring temporal dependencies and extracting meaningful information. Key concepts include trend analysis, seasonality detection, and the identification of irregular fluctuations.

One crucial aspect of time series analysis is seasonal decomposition, which disentangles a time series into its constituent components: trend, seasonality, and residual (noise). This process enables the extraction of underlying patterns and their variations over time. In the context of Bloemfontein wheat farms, understanding the seasonal patterns of climate variables and their impact on wheat yield can provide valuable insights for crop management.

2.1.1. ARIMA

Models are a staple in time series analysis, particularly for capturing non-seasonal trends and irregular fluctuations. These models encompass autoregressive (AR) terms that account for the dependence on past observations, integrated (I) terms to address non-stationarity, and moving average (MA) terms to model residual variations. ARIMA models provide a valuable tool for understanding the time-dependent relationships between climate variables and wheat yield.

2.1.2. Seasonal ARIMA (SARIMA)

Seasonal ARIMA (SARIMA) models extend the ARIMA framework to account for seasonality for data exhibiting seasonal patterns, such as climate variables impacted by yearly cycles. These models include additional terms to capture seasonal dependencies, making them particularly relevant for our study's focus on Bloemfontein's wheat yield data influenced by climate variability and annual agricultural cycles.

2.1.3. Long Short-Term Memory (LSTM)

Modern advancements in machine learning have introduced LSTM networks, which are well-suited for capturing complex temporal dependencies in time series data. LSTMs are a recurrent neural network that can model long-range interactions and capture intricate patterns. Their application to agricultural data offers the potential to uncover nonlinear relationships between climate variables and wheat yield. Time series analysis methodologies provide a systematic approach to unravel the dynamics of climate variability's impact on wheat yield. By applying techniques such as ARIMA, SARIMA, and LSTM to the available data, we aim to derive actionable insights that contribute to understanding the multifaceted interactions between climate patterns and agricultural productivity.

Time series analysis methods offer a structured approach to deciphering the intricate temporal dynamics of climate variables and wheat yield in the context of Bloemfontein wheat farms. The subsequent sections of our study will apply these techniques to the data, providing a deeper understanding of climate variability's implications for local agriculture.

2.2. Considerations for Bloemfontein Wheat Farms

As we delve deeper into the investigation of climate variability's impact on wheat yield in Bloemfontein wheat farms, it is essential to consider various factors that significantly shape the region's agricultural landscape. These considerations encompass the area's unique characteristics and the challenges farmers face in adapting to changing climate patterns.

2.2.1. Semi-Arid Climate

Bloemfontein's semi-arid climate introduces distinct challenges for wheat farming. With limited and erratic rainfall, water availability is a critical factor influencing crop growth and yield. Effective water management strategies, including efficient irrigation techniques and water conservation measures, are imperative for mitigating the risks associated with drought and water stress during crucial growth stages [1, 8].

2.2.2. Soil Health and Fertility

The health and fertility of soils in Bloemfontein wheat farms significantly impact crop productivity. Soil characteristics, such as texture, organic matter content, and nutrient levels, influence wheat's ability to access essential nutrients and moisture. Maintaining soil health through proper management practices, including crop rotation and nutrient supplementation, is crucial for sustaining optimal yields [3, 6].

2.2.3. Crop Varieties and Resilience

The choice of wheat varieties can significantly influence a farm's resilience to climate variability. Selecting climate-resilient and drought-tolerant wheat cultivars ensures stable yields under fluctuating climatic conditions. Furthermore, diversifying crop varieties can enhance a farm's adaptability to changing climate patterns and reduce vulnerability to yield losses [4, 7].

2.2.4. Adaptive Agricultural Practices

Farmers in Bloemfontein must adopt adaptive agricultural practices that align with the region's climate patterns. Adjusting planting and harvesting schedules, implementing water-efficient irrigation methods, and optimizing fertilizer application can mitigate the effects of climate variability on wheat yield. Adaptive practices also include monitoring weather forecasts and adjusting management strategies accordingly [5, 11].

2.2.5. Integrated Pest Management

Managing pests and diseases is vital for preserving crop health and yield. Bloemfontein's climate variability can impact pest lifecycles and disease prevalence. Implementing integrated pest management strategies that combine biological, cultural, and chemical control methods can minimize pest and disease-related yield losses [12].

2.2.6. Community Engagement and Knowledge Sharing

Collaboration and knowledge sharing within the agricultural community are crucial in building resilience to climate variability. Farmers, researchers, and agricultural extension services can collaborate to disseminate climate-smart practices, share experiences, and develop localized strategies for mitigating climate-related risks.

2.2.7. Policy and Support

Government policies, subsidies, and support programs can facilitate the adoption of climate-resilient practices among wheat farmers. Policy initiatives that promote sustainable water management, provide access to advanced agricultural technologies and offer financial incentives can enhance the resilience of Bloemfontein wheat farms. Understanding these considerations is integral to the success of our investigation. By recognizing the unique challenges Bloemfontein wheat farms face and the strategies employed to address them, we can contextualize our analysis and develop practical insights for local farmers. The subsequent sections of our study will incorporate these considerations as we analyze climate and yield data [13].

These considerations provide a holistic view of the various aspects that need to be considered when investigating the impact of climate variability on wheat yield in Bloemfontein wheat farms. Integrate these considerations into your study to provide a comprehensive understanding of the challenges and opportunities in the region's agricultural landscape.

2.3. Data Selection and Understanding

The climate data for this study was sourced from Meteostat, a comprehensive weather database renowned for its accurate and reliable historical climate information. The dataset encompasses a range of climatic variables crucial for analyzing the impact of climate variability on wheat yield in Bloemfontein. These variables include temperature (minimum and maximum), precipitation, humidity, and other relevant parameters. Meteostat compiles weather information from various weather stations located in and around Bloemfontein. These stations record meteorological parameters at regular intervals, typically hourly or daily, providing a comprehensive temporal perspective on climate variability. Data was collected for an extended period to capture long-term trends, anomalies, and potential cyclical patterns that may influence wheat yield. Wheat yield information was obtained from the Department of Agriculture in the Free State province, providing accurate and comprehensive records of wheat productivity in the Bloemfontein region. The yield data encompass a series of yearly records detailing the amount of wheat harvested in metric tons per hectare. The Department of Agriculture's records were obtained through formal requests and collaboration. The dataset contains historical data on wheat yield for a substantial period, aligning with the temporal scope of the climate data. The yield records are crucial for establishing correlations between climate variability and wheat yield fluctuations. Understanding the complexities and intricacies of the climate and yield data is pivotal to uncovering meaningful insights. Detailed statistical and exploratory analyses will be conducted to identify trends, anomalies, and potential relationships. Correlation analyses between climate variables and wheat yield will be employed to quantify the degree of impact that climate variability exerts on agricultural productivity.

Preparing and converting data are basic steps before data modeling [14]. Clean data is produced as a consequence of data pre-processing and transformation [15]. In Python, the following data preprocessing and transformation procedures will be applied. First, locate and handle missing values and outliers. The data will then be converted using standardization in the second phase to guarantee that each observation is on the same scale. Rigorous preprocessing and quality control measures were applied to ensure the reliability and consistency of the climate data. This involved removing outliers, filling missing data points through interpolation, and cross-validating records against neighboring stations. Data cleaning procedures were carried out meticulously to minimize the impact of data gaps or inconsistencies on subsequent analyses.

2.4. Modelling

The Prepared and converted data will be analyzed using a regression algorithm, namely Multiple Linear Regression and time series algorithms such as ARIMA, SARIMA, vector Auto-Regressive (VAR), Long Short Term Memory (LSTM), and Facebook Prophet. Autoregressive Integrated Moving Average (ARIMA) and Seasonal ARIMA (SARIMA) models are widely used in time series analysis to capture temporal patterns and forecast future values. These models are well-suited for understanding the relationships between climate variability and wheat yield fluctuations. Facebook Prophet is a robust forecasting tool that handles seasonal effects, holidays, and trend changes. This model will allow us to capture the climate variability's impact on wheat yield and any additional calendar-related patterns. The Vector Autoregression (VAR) model is particularly useful for examining the dynamic interactions between multiple time series variables. In this case, we can analyze how different climate variables interact with each other and collectively influence wheat yield. Long Short-Term Memory (LSTM) networks are powerful tools for handling sequences and patterns in time series data. Applying an LSTM model to climate and yield data can unveil complex nonlinear relationships that traditional methods might miss. Multiple Linear Regression is a baseline model for assessing the direct linear relationships between climate variables and wheat yield. By considering multiple predictor variables, we can quantify the individual contributions of different climate factors to yield fluctuations.[7]. The data structure and model flow diagrams are shown in Figure 1 and Figure 2 As shown in Figure 1, the dataset consists of two sources: the weather dataset and wheat production data. Figure 2 shows the flow of the project and how it will be compiled using the programming language Python.

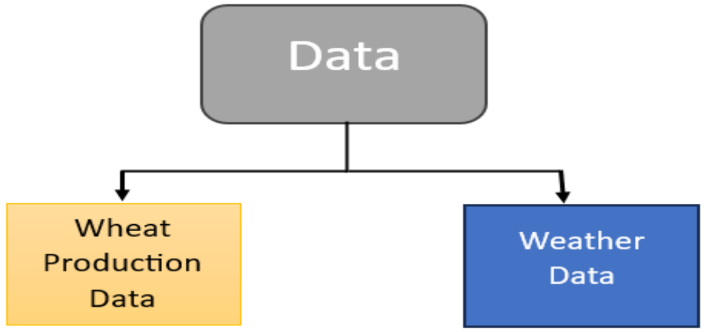

Figure 1.
Data structure.

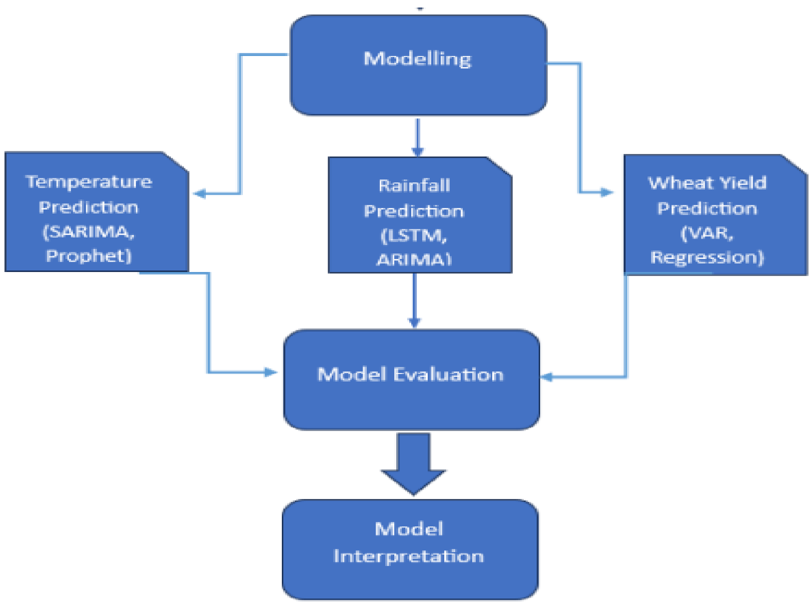

Figure 2.
Model flow diagram.

2.4.1. ARIMA

In ARIMA, the abbreviation "AR" stands for "autoregressive," meaning the model relies on the dependent relationship between an observation and its previous values. The number of preceding inputs used to predict the next value is called order and is usually referred to as p. ARIMA models are based on the idea that past values of a time series can be used to predict future value. The model has three main components:

- Autoregression (AR): The AR component uses past time series values to predict future values. It is called auto-regression because it regresses the variable against itself.
- Integrated (I): The "I" in ARIMA represents the difference in the time series to achieve stationarity. The time series' statistical properties, such as its mean, variance, and auto-correlation, remain constant over time.

ISSN: 2576-8484
Vol. 8, No. 5: 1213-1234, 2024
DOI: 10.55214/25768484.v8i5.1826

- Moving Average (MA): The MA component uses past errors (the difference between actual and predicted values) to predict future values.

The general formula for an ARIMA(p, d, q) model is shown in Equation 1:

$y(t) = c + \phi(1)y(t-1) + \ldots + \phi(p)y(t-p) + e(t) + \theta(1)e(t-1) + \ldots + \theta(q)e(t-q)$ **Equation 1**

Where:

- $y(t)$ is the value of the time series at time t
- c is a constant term
- $\phi(1), \ldots, \phi(p)$ are the AR coefficients of the model, with p denoting the number of lags used in the auto-regression
- $e(t)$ is the error term at time t
- $\theta(1), \ldots, \theta(q)$ are the MA coefficients of the model, with q denoting the number of lags used in the moving average
- d is the degree of differencing needed to make the time series stationary

To fit an ARIMA model, you would typically estimate the values of p, d, and q based on the properties of the chronological data under examination using statistical tests or visual inspection. Once you have estimated these values, you can estimate the model parameters (phi and theta coefficients) using maximum likelihood estimation or another optimization technique. Some of the algorithms, such as mean squared error, root mean squared error, and Akaike information criterion, can be used to evaluate the performance of an ARIMA model. These measures indicate how well the model fits the observed data and can be generalized to new data.

2.4.2. Regression Model

Regression models are widely used in analyzing the relationship between dependent variables and one or more independent variables. They can be applied to evaluate the impact of climate change on crop production by examining the dependence of crop yield on climate variables.

- Multiple Linear Regression (MLR) is a statistical method that models the intricate relationship between a dependent variable and two or more independent variables. This extends the fundamental concept of simple linear regression, which examines the connection between two variables, to scenarios where the outcome's variability is influenced by multiple predictors simultaneously. Multiple Linear Regression will be employed to assess the impact of various climate variables on wheat yield.

The Multiple Linear Regression model can be represented Equation 2:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon \quad (2)$$

Where:

- y is the dependent variable.
- $x_1, x_2, \ldots. x_p$ are the independent variables
- $\beta_0, \beta_1, \beta_2, \beta_p$ The coefficients represent the relationship between the independent and dependent variables.
- var epsilon is the error term, representing the unexplained variability.

The goodness of fit of the Multiple Linear Regression model can be evaluated using metrics like the coefficient of determination (R^2) and the adjusted R^2 . These metrics indicate the proportion of variance in the dependent variable explained by the independent variables.

2.4.3. Facebook Prophet

Facebook Prophet is a versatile time series forecasting tool developed by Facebook's Core Data Science team. It is specifically designed to handle a wide range of time series data, making it particularly valuable for capturing climate-driven variations in wheat yield. Prophet is equipped with several features that make it suitable for your analysis:

- Flexible Handling of Seasonality: Prophet can effectively model various seasonal patterns, including daily, weekly, and yearly cycles, as well as holidays that influence yield patterns.
- Accounting for Holidays: The model can account for holidays or significant events that impact wheat yield. This is particularly useful in capturing climate-related holidays like heatwaves or periods of water scarcity.
- Trend Modeling: Prophet captures both short-term and long-term trends, allowing you to assess the impact of climate variability while considering the broader context of yield fluctuations.
- Automatic Handling of Missing Data: Prophet can automatically handle missing data points and outliers, reducing the need for extensive data preprocessing.

2.4.4. Vector Autoregression (VAR)

Vector Autoregression (VAR) is a statistical method used to analyze the relationships between multiple time series variables. In the study, VAR will be employed to explore how different climate variables interact with each other and collectively influence wheat yield. VAR extends the concept of univariate autoregression to multiple variables. Instead of modeling each variable separately, VAR models the joint behavior of all variables in the system. This approach is particularly valuable when examining how temperature, precipitation, and other climate factors interact and contribute to yield fluctuations. VAR is based on the assumption that each variable in the system is influenced by its past values and the past values of all other variables. This captures the dynamic relationships and potential feedback loops between climate variables and wheat yield. By considering these interdependencies, VAR offers a more comprehensive understanding of the system's behavior. Selecting the appropriate lag order is a crucial step in VAR modeling. The lag order determines how many previous time steps are used to predict the current values of the variables. Techniques like the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) can guide the selection of an optimal lag order that balances model complexity and predictive accuracy. Interpreting VAR outputs involves analyzing the coefficients of lagged variables in the system. Positive or negative coefficients indicate the direction and magnitude of the influence between variables. Additionally, analyzing the IRFs and FEVD plots provides insights into the dynamic interactions and the relative importance of climate variables in affecting wheat yield.

2.4.5. Seasonal Autoregressive Integrated Moving Average (SARIMA)

SARIMA is an extension of the ARIMA model that incorporates seasonal components. It considers the autoregressive and moving average components and the seasonal differences in the data. This makes SARIMA well-suited for capturing the seasonal variations driven by climate factors. SARIMA models consist of three main components: autoregressive (AR), integrated (I), and moving average (MA). The seasonal aspect introduces additional parameters denoted as P, D, and Q, representing the seasonal autoregressive, integrated, and moving average components represented in Equation 3.

$$(1 - \phi_1 B - \phi_1 B^s)(1 - B)^d (1 - B^s)^D y_t = (1 + \theta_1 B + \Theta_1 B^s)(1 - B)^d (1 - B^s)^D \varepsilon_t \quad (3)$$

Selecting the appropriate order of p, d, q, P, D, and Q involves analyzing the autocorrelation and partial autocorrelation plots. Additionally, the decomposition of the time series into its seasonal, trend, and residual components aids in understanding the data characteristics and seasonality. Once the SARIMA model is fitted to the data, it can be used for forecasting. The model captures the seasonal patterns driven by climate variability and the temporal dynamics influencing wheat yield. Forecasts generated by the SARIMA model provide insights into how climate fluctuations affect yield over different periods. Interpreting SARIMA outputs involves analyzing the coefficients of the autoregressive and moving average terms and the seasonal components. The fitted model's residuals should be examined for autocorrelation and normality to ensure the model adequately captures the underlying patterns.

2.4.6. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) neural networks are a specialized type of recurrent neural network (RNN) designed to capture long-range dependencies and patterns in time series data. LSTM networks are particularly effective in capturing temporal relationships in sequences of data. Unlike traditional feedforward neural networks, LSTMs have internal memory cells and gating mechanisms, allowing them to store and retrieve information selectively over long intervals. LSTM models consist of layers of LSTM units, each with multiple memory cells and gates. Proper tuning of hyperparameters is essential for optimal model performance. LSTMs inherently account for sequential patterns in the data, which is crucial for capturing the dynamics of climate variability and their influence on wheat yield. The network learns to recognize short-term fluctuations, long-term trends, and recurring patterns driven by climate factors. LSTM networks consist of multiple memory cells (or units) that have gating mechanisms to control the flow of information. These gates include the forget gate, input gate, and output gate, which work together to manage the cell's state and output. Here's an overview of the key equations shown in Equations 4-10:

i. Forget Gate: This gate determines which information from the cell's previous state should be discarded.
   i. $f_t = \sigma(W_f * \left[h_{t-1,} x_t\right] + b_f$ (4)
ii. Input Gate: This gate decides which new information should be stored in the cell's state.
   i. $i_t = \sigma(W_i * \left[h_{t-1,} x_t\right] + b_i)$ (5)
iii. Candidate New State: This equation calculates the candidate values that could be added to the cell's state.
   i. $\tilde{C}_t = \tanh(W_C * \left[h_{t-1,} x_t\right] + b_C$ (6)
iv. Cell State Update: The current cell state is updated by combining the results from the forget gate, input gate, and candidate new state.
   i. $C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$ (8)
v. Output Gate: This gate determines the output based on the updated cell state.
   i. $o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o)$ (9)
vi. Hidden State Update: The current hidden state is calculated by passing the updated cell state through the output gate.
   i. $h_t = o_t * \tanh(C_t)$ (10)

3. Results

In this pivotal section, we embark on a journey through the empirical outcomes of the study. The diligent application of advanced modeling techniques has enabled us to forecast temperature patterns, rainfall variations, and wheat production levels within the dynamic agricultural landscape of Bloemfontein. The cornerstone of the study was the utilization of cutting-edge predictive models, each tailored to address a specific facet of the complex relationship between climate variables and wheat production. Our arsenal included Facebook Prophet and SARIMA for temperature forecasts, ARIMA and LSTM for rainfall predictions, and VAR and Multiple Linear Regression models for wheat production estimations.

In this comprehensive presentation of results, we will systematically dissect the outcomes of each modeling endeavor. From the fluctuations in temperature and precipitation to the intricacies of wheat yield forecasts, we delve into the empirical data discerningly. As we navigate this terrain, we aim to decipher the implications of our findings, particularly in the context of climate change's inevitable influence on agricultural sustainability. To gain insight into the evolving temperature dynamics within Bloemfontein, we harnessed the predictive capabilities of two robust models: Facebook Prophet and SARIMA. This section provides a comprehensive overview of our temperature prediction results, offering a lens through which to view the precision and nuances of our models. TABLE 1 and TABLE 2

ISSN: 2576-8484
Vol. 8, No. 5: 1213-1234, 2024
DOI: 10.55214/25768484.v8i5.1826

present the component of the SARIMA model result for the average temperature. The interpretations of these components is covered below.

3.1. Temperature
3.1.1. SARIMA Results

Table 1.
SARIMA model results for temperature (tavg).

Component	Coefficient	Standard Error	Z-statistic	P-value
ar.L1	0.0121	0.015	0.811	0.417
ar.L2	0.9859	0.015	65.783	0.000
ma.L1	0.2720	0.057	4.779	0.000
ma.L2	-0.8864	0.030	-29.740	0.000
ma.L3	-0.1614	0.056	-2.890	0.004
ar.S.L12	0.9963	0.003	378.319	0.000
ma.S.L12	-0.7816	0.065	-12.090	0.000
ma.S.L24	-0.0226	0.068	-0.335	0.738

Table 2.
SARIMA model results for temperature (tavg).

Component	Value
Dep. Variable	tavg
No. Observations	361
Model	SARIMAX(2, 0, 3)x(1, 0, [1, 2], 12)
Log Likelihood	-660.594
Date	Mon, 07 Aug 2023
AIC	1339.188
Time	15:39:50
BIC	1374.188
Sample	01-31-1990 - 01-31-2020
HQIC	1353.103

3.2. The Interpretations of the Results for Table 1 and Table 2:

i. Dep. Variable (Dependent Variable: tavg): This indicates that the analysis is focused on the variable "tavg," which represents average temperature.
ii. No. Observations (Number of Observations: 361): This shows that 361 data points or observations are included in the analysis. These observations represent measurements of the variable (average temperature) taken over time.
iii. Model (SARIMAX(2, 0, 3)x(1, 0, [1, 2], 12)): This indicates the specific SARIMA model that has been applied to the data. SARIMA stands for Seasonal AutoRegressive Integrated Moving Average. The numbers in parentheses represent the model's order:
    a. (2, 0, 3): Represents the non-seasonal part of the SARIMA model.
    b. (1, 0, [1, 2], 12): Represents the seasonal part of the model, with a seasonal order of (1, 0, [1, 2]) and a seasonal period of 12 (likely indicating a yearly seasonality).

iv. Log Likelihood (-660.594): The log likelihood measures how well the model fits the data. A higher log-likelihood value indicates a better fit. In this case, the negative value suggests that the model provides a better fit than a null model with no predictors.

v. AIC (Akaike Information Criterion: 1339.188): The AIC measures the model's goodness of fit while penalizing for complexity. Lower AIC values indicate a better trade-off between model fit and complexity. An AIC of 1339.188 suggests that the SARIMA model provides an excellent fit to the data while considering its complexity.

vi. BIC (Bayesian Information Criterion: 1374.188): Similar to AIC, the BIC also measures model fit while penalizing for complexity. Like the AIC, lower BIC values suggest a better trade-off. A BIC value of 1374.188 indicates that the SARIMA model reasonably fits the data while penalizing complexity.

vii. Sample (01-31-1990 - 01-31-2020): This indicates the period covered by the analysis, from January 31, 1990, to January 31, 2020.

viii. HQIC (Hannan-Quinn Information Criterion: 1353.103): The HQIC is another model selection criterion considering goodness of fit and complexity. It can help in comparing models, and lower values are preferred.

Our journey into temperature forecasting begins with a quantitative assessment of our models' performance. To gauge the accuracy of our predictions, we turn to key performance metrics, namely the MSE and RMSE. The SARIMA model achieved an MSE of 7.36 and RMSE of 2.71; the lower the MSE and RMSE, the more reliable the model is. The Monthly Predictions and Yearly Predictions from the SARIMA model are shown in Figure 3 and Figure 4. Table 3 shows the temperature data for between the year 2022 and 2023 recorded on the last day of each month. Looking at months such as June and July this checks out as during those months, it is winter in South Africa so the temperatures tend to be very low.

Table 3.
Temperature data for 2022-2023.

Date	Temperature
2022-12-31	23.649303
2023-01-31	23.977977
2023-02-28	23.495615
2023-03-31	21.369650
2023-04-30	17.027166
2023-05-31	13.273785
2023-06-30	9.952692
2023-07-31	9.869566
2023-08-31	12.410708
2023-09-30	16.977953
2023-10-31	20.279060
2023-11-30	21.981281

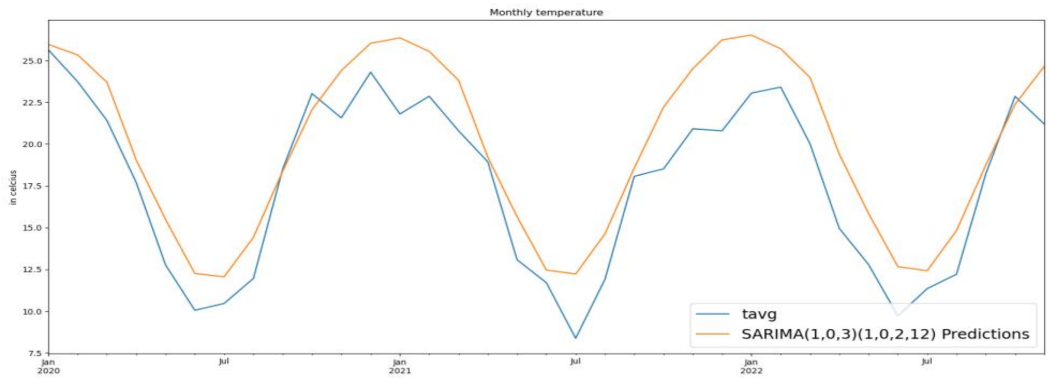

Figure 3.
Monthly temperature predictions.

3.1.2. Facebook Prophet Results

The Facebook Prophet model achieved stellar results in temperature forecasting, achieving a Root Mean Squared Error of 3.54 and a Mean Squared Error of 12.51. These values depict that the model is also reliable. Yearly Temperature Predictions and the actual temperature versus the predicted from the Facebook Prophet model are shown in Figure 3 and Figure 4. TABLE 4 shows Facebook Prophet temperature forecast results and the terms ds, yhat, yhat_lower, and yhat_upper are components of the output from its predictive model.

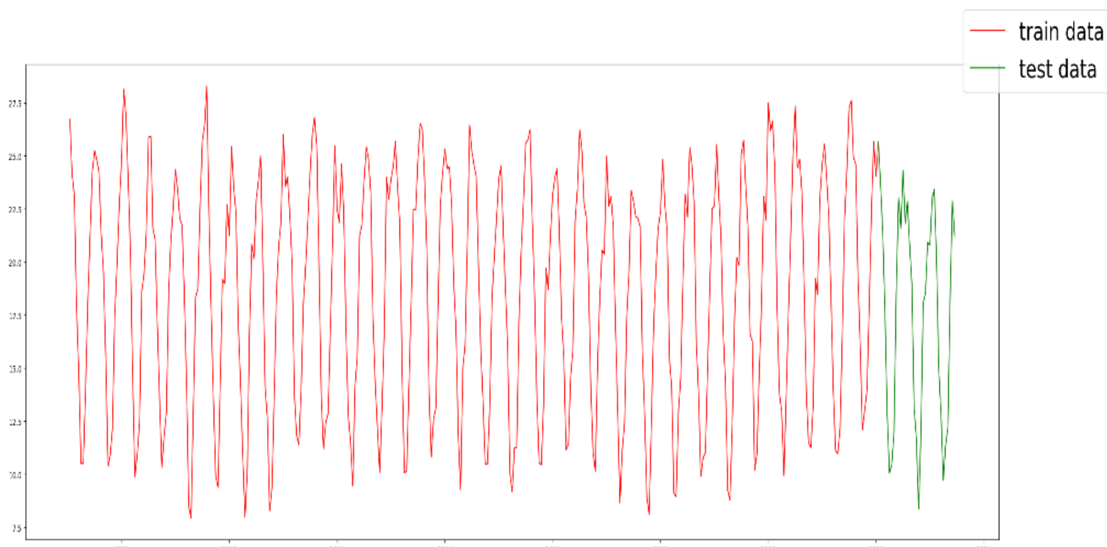

Figure 4.
Yearly temperature predictions.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1213-1234, 2024
DOI: 10.55214/25768484.v8i5.1826

Table 4.
Facebook prophet forecast results.

ds	yhat	yhat_lower	yhat_upper
2023-01-31	24.294721	19.720592	29.121060
2023-02-28	22.815139	18.232665	27.218555
2023-03-31	19.526470	15.142023	23.981954
2023-04-30	14.800707	10.304158	19.407743
2023-05-31	10.855844	6.534459	15.261253
2023-06-30	9.424365	5.056646	14.118190
2023-07-31	10.486375	6.138765	15.120995
2023-08-31	14.512881	9.780269	19.193929
2023-09-30	18.190752	13.779352	22.705058
2023-10-31	20.553815	16.045287	25.070549
2023-11-30	22.384519	17.829002	26.966059
2023-12-31	24.191557	19.640244	28.821561

From Table 4 ds, represents the date column in the input DataFrame that Prophet uses to make its forecasts, yhat is the predicted value, yhat_lower and yhat_upper represent the lower and upper bounds of the uncertainty interval for the forecast, respectively. They provide a range within which the actual value is expected to lie with a certain probability. When we look at June and July we can see the temperatures are also low as this is a winter period. The yhat_lower and yhat_upper provide reliable results.

These visual aids unveil the cyclical patterns and seasonal variations within the temperature data and provide a canvas upon which deviations between predictions and observations become evident. In these deviations, we uncover the challenges and opportunities presented by climate change, as reflected in the dynamic temperature landscape of Bloemfontein. In conclusion, the SARIMA model is way more effective than Facebook Prophet regarding temperature forecasting based on the RMSE and MSE.

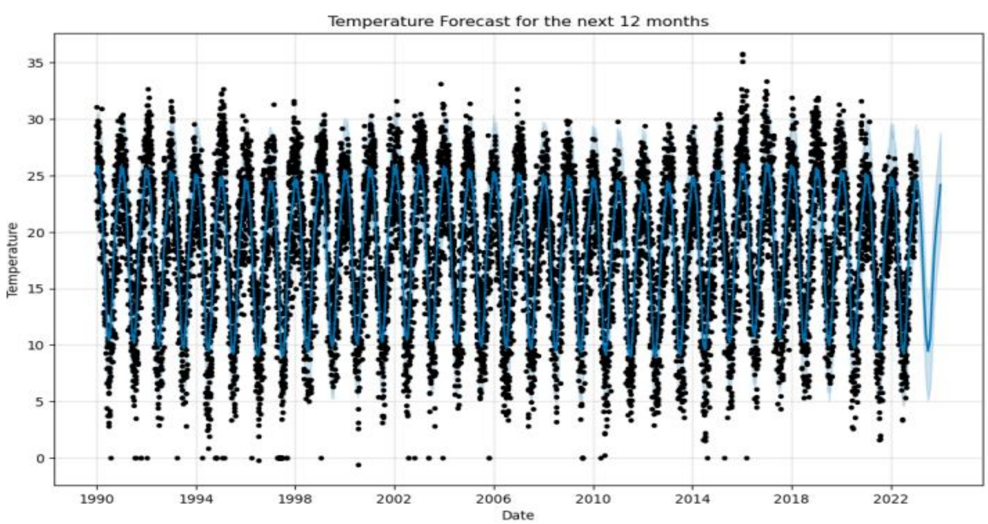

Figure 5.
Yearly temperature predictions.

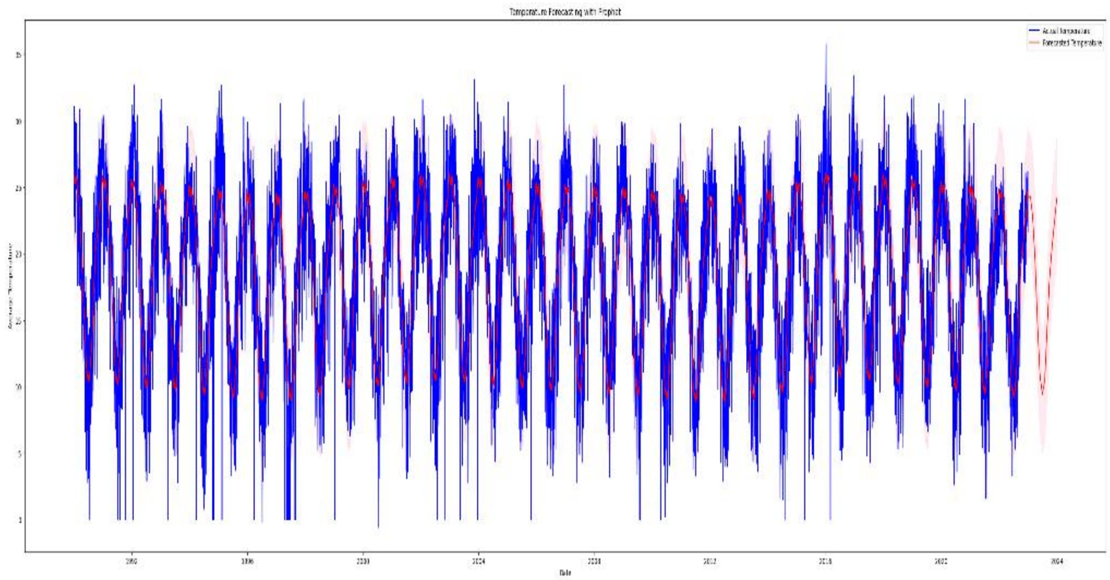

Figure 6.
Actual vs predicted temperature.

3.2. Precipitation
3.2.1. ARIMA Results

To illuminate the intricacies of precipitation variations and their implications for wheat production, we harnessed the power of two distinct forecasting approaches: ARIMA and LSTM. The ARIMA model achieved an RMSE of 46.6 and an MSE of 2173.1. These may seem high, but they are perfectly normal regarding precipitation. Yearly precipitation predictions from the ARIMA model are shown in Figure 7.

Table 5.
ADF test results.

Statistic	Value
ADF	-5.741
p-value	6.275×10^{-7}
Critical Values (1%)	-3.447
Critical Values (5%)	-2.869
Critical Values (10%)	-2.571

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 5: 1213-1234, 2024
DOI: 10.55214/25768484.v8i5.1826

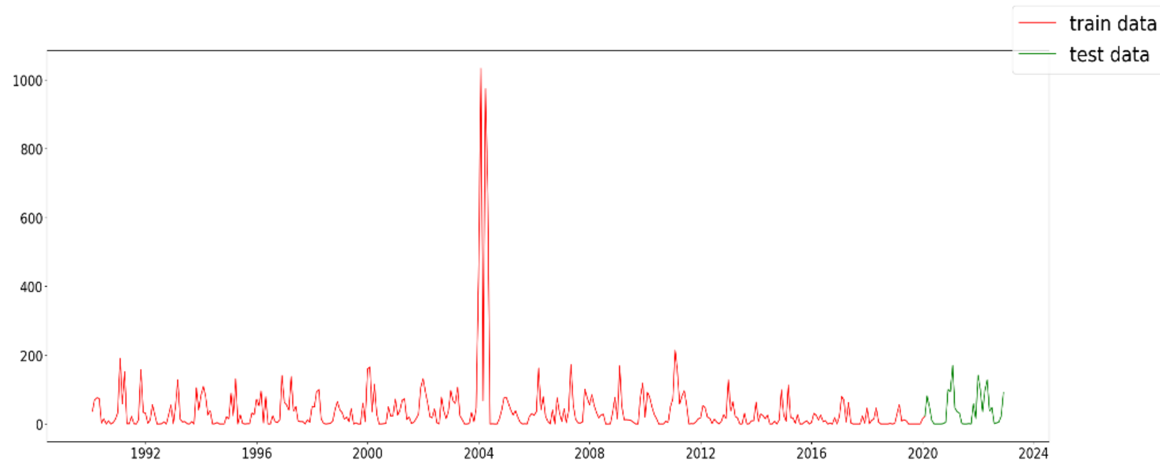

Figure 7.
Yearly precipitations predictions.

The findings presented in TABLE 5 stem from the Augmented Dickey-Fuller (ADF) test, a widely utilized tool in time series analysis for detecting the presence of a unit root within a time series dataset. This test is crucial in determining whether a given time series exhibits characteristics of stationarity or nonstationarity. To shed light on the significance of the ADF test results, let us delve into an explanation of its fundamental components from TABLE 5:

i. ADF Statistic: -5.741029316952197. The ADF statistic is a test statistic computed during the ADF test. It represents how much the time series needs to be differenced (i.e., how many lag differences are required) to become stationary. The ADF statistic is -5.741, which is highly negative. This suggests that the time series is likely stationary, as a more negative ADF statistic indicates stronger evidence against the presence of a unit root.
ii. P-value: 6.275099685404499e-07. The p-value is associated with the ADF statistic and is used to assess the statistical significance of the test. It measures the probability of obtaining an ADF statistic as extreme as the one observed if the null hypothesis were true (null hypothesis: the time series has a unit root, i.e., is non-stationary). The extremely low p-value (approximately 6.275e-07, which is close to zero) suggests strong evidence against the null hypothesis of a unit root. This further supports the conclusion that the time series is stationary.
iii. Since the ADF statistic (-5.741) is significantly more negative than these critical values, it suggests strong evidence against the presence of a unit root, indicating that the time series is likely stationary.

3.2.2. LSTM Results

The LSTM model was trained for 400 epochs using the sequential model, one Density layer, and the Adam optimizer. The model achieved an RMSE of 41.06 and MSE of 1942.2. Again, these may seem high, but in the case of precipitation prediction, they are normal. Yearly precipitation predictions and the monthly precipitation prediction from the LSTM model are shown in Figures 8 and 9.

DOI: 10.55214/25768484.v8i5.1826

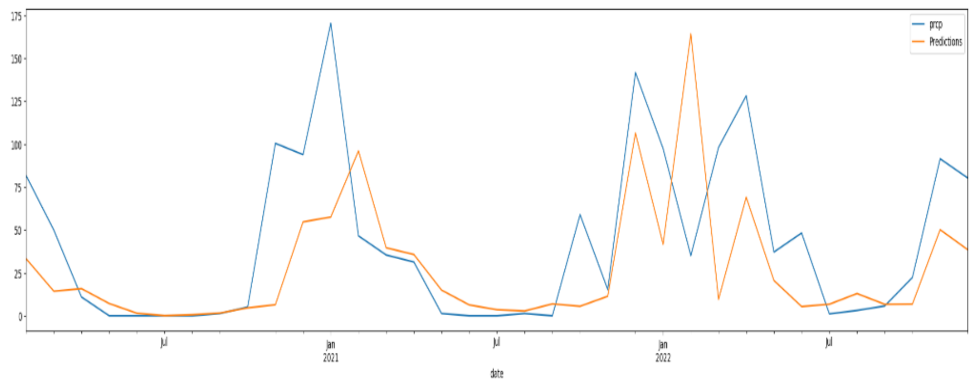

Figure 8.
Yearly precipitation predictions.

One study [16] encompassed an in-depth analysis of monthly precipitation data, spanning from 1967 to 2017, and an examination of annual precipitation using Grey Theory methods. Additionally, they harnessed advanced modeling techniques, including Wavelet Transformation, ARIMA, and LSTM, to unravel the intricacies of these time series. The outcomes of these analyses are summarized below in TABLE 6 which shows the precipitation prediction results for RMSE in three different station using the ARIMA and LSTM models:

Table 6.
Precipitation prediction RMSE.

Station	Model	RMSE (mm)
Changchun	ARIMA	38.698
	LSTM	34.571
Linjiang	ARIMA	42.739
	LSTM	38.994
Qian Gorlos	ARIMA	37.535
	LSTM	34.509

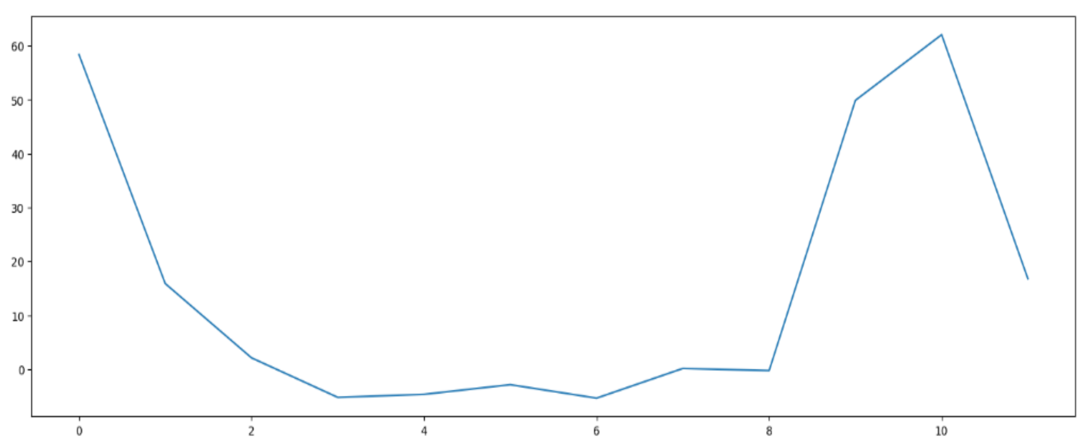

Figure 9.
Monthly precipitations prediction.

ISSN: 2576-8484
Vol. 8, No. 5: 1213-1234, 2024
DOI: 10.55214/25768484.v8i5.1826

3.3. Wheat Production

3.3.1. Vector Autoregression

Armed with the formidable Vector Autoregression (VAR) and Multiple Linear Regression tools, we ventured into wheat production forecasting. These meticulously crafted and fine-tuned models were tasked with unraveling the intricate web of factors that govern wheat yields in the face of shifting climatic patterns. First, we looked at the relationship between temperature and precipitations to determine whether they affect one another. That relationship is shown in Figure 10. For the VAR model, it achieved an MSE of 12.56 and an RMSE of 16.7. These numbers are not low thus, this shows the model is not that reliable.

Table 7.
VAR model results.

Order	AIC	BIC	FPE	HQIC
0	40.32	40.47	3.248e+17	40.36
1	-10.26*	-9.676*	3.517e-05*	-10.11*
2	-9.561	-8.530	7.421e-05	-9.287
3	-7.830	-6.358	0.0004	-7.439
4	-9.877	-7.962	7.578e-05	-9.369
5	-6.768	-4.412	0.0036	-6.143
6	-5.996	-3.198	0.01370	-5.253

Table 7 shows the results of different VAR (Vector Auto Regressive) model specifications, with each row corresponding to a different model order and here is the interpretation of the table.

i. Order: This column indicates the order in which the VAR model is considered. In time series analysis, the order of a VAR model represents how many lagged values of each variable are included. For example, a VAR(1) model has only one lag, while a VAR(2) model consists of two lags.
ii. AIC (Akaike Information Criterion): AIC measures a model's goodness of fit while penalizing for model complexity. Lower AIC values are better, as they indicate a better trade-off between model fit and complexity. In the table, some models have negative AIC values, suggesting they are likely suitable fits for the data. The asterisk (*) next to -10.26 indicates this model has the lowest AIC among the ones presented.
iii. BIC (Bayesian Information Criterion): Similar to AIC, BIC measures model fit while penalizing complexity. Like AIC, lower BIC values are preferred. On the table, the models with lower BIC values, denoted by asterisks (*), indicate a better trade-off between model fit and complexity.
iv. FPE (Final Prediction Error): The FPE measures the model's prediction error. It is used in model selection, but unlike AIC and BIC, there is no clear guideline regarding the magnitude of FPE values. Typically, lower FPE values are considered better.
v. HQIC (Hannan-Quinn Information Criterion) is another model selection criterion that, like AIC and BIC, considers goodness of fit and model complexity. Lower HQIC values are preferred. Models with asterisks (*) suggest a better trade-off between model fit and complexity.

Based on these results, The model with order 1 appears to have the lowest AIC, BIC, and HQIC values, indicating that it is a good fit for the data while being relatively simple. This model is marked with asterisks. As the order of the VAR model increases beyond 1, the AIC, BIC, and HQIC values generally increase, suggesting that the models become less favorable regarding model fit and complexity tradeoff. The results indicate that a VAR(1) model performs well according to AIC, BIC, and HQIC criteria.

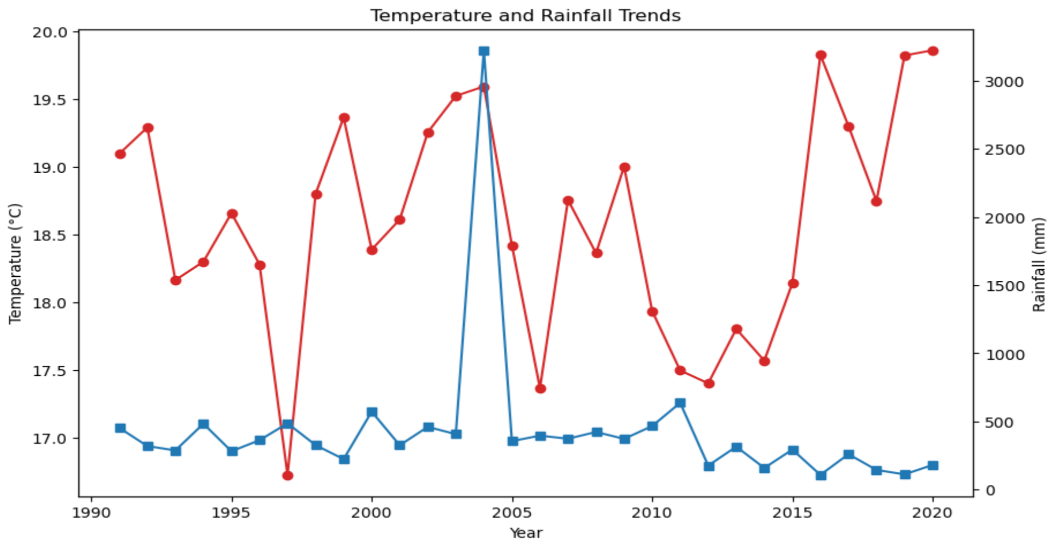

Figure 10.
Temperature and precipitation trend.

In addition, certain variables were explored, including the production yield and the area, production and precipitations- and then production and temperature. Temperature and Production Trend and Production + Area and Production + Precipitation Trend are shown in Figure 11 and Figure 12.

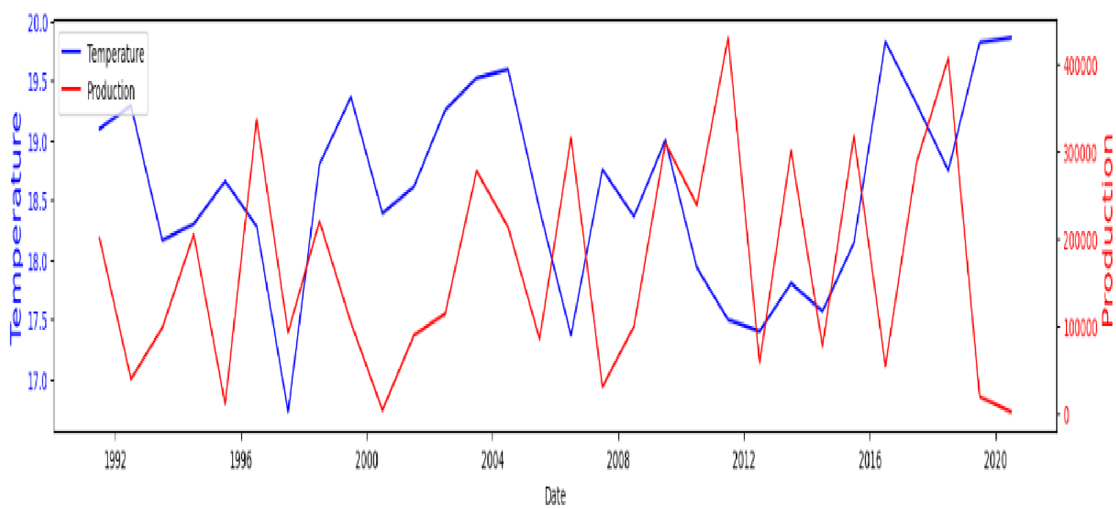

Figure 11.
Temperature and production trend.

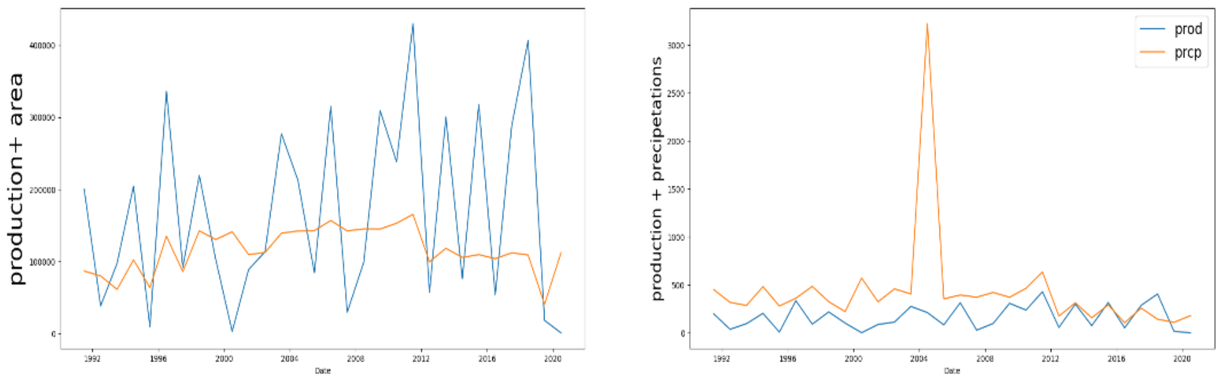

Figure 12.
Production + area and production + precipitation trend.

3.3.2. Multiple Linear Regression

The independent variables for the multiple linear regression are precipitation and production, while the dependent variable is temperature. The model was trained, and from the evaluation, the MSE was 81613.5, and the RMSE was 89808.4. This shows that the multiple linear regression performed less better than VAR.

Table 8.
Wheat yield prediction - testing (2021).

Year	Predicted Yield	Real Yield	Precipitation (mm)	Temperature (°C)
2010	175613.98	238312.04	465.9	17.9
2021	111491.434505	143380.34	180	19.5

Table 8 compares wheat yield predictions and actual real yields for two different years, 2010 and 2021. The model is not far off, as seen from the predicted wheat yield and the real values. In conclusion, six models were explored, and two models were allocated for each task. In terms of Temperature forecasting, SARIMA, and Facebook Prophet were used, and from the results, SARIMA performed better than Prophet as it could make reliable predictions. For Precipitation/Rainfall prediction, ARIMA and LSTM were explored, and LSTM was more reliable than ARIMA. So, for rainfall forecasting, LSTM was the model that produced better results. Lastly, for Wheat Production Yield, two models were utilized: Multiple linear regression and VAR. In this case, VAR performed better as the MSE and RMSE values are reliable. The implications of our wheat production predictions extend beyond data analysis. They reach the heart of Bloemfontein's farming community and the broader agricultural landscape. These results can potentially guide decision-making processes, shape policies, and inform adaptive strategies in a changing climate. As we traverse this landscape of data and agriculture, we consider the resilience and adaptability of local farmers and stakeholders. The results serve as a compass, pointing toward potential challenges and opportunities in sustaining wheat cultivation amidst climatic uncertainties. They underscore the importance of proactive measures, from crop diversification to resource management, in safeguarding agricultural livelihoods.

4. Discussion

The study's initial step involved a meticulous analysis of historical climate data that included temperature and precipitation. In the first component, the project undertook the task of temperature forecasting in the Bloemfontein region. Applying two distinct time series forecasting techniques, SARIMA (Seasonal AutoRegressive Integrated Moving Average) and Prophet, achieved this. SARIMA, known for capturing seasonal patterns, was deployed to model the temporal temperature variations. Meanwhile, the Prophet provided a complementary perspective with its capacity to handle holiday effects and special events. Utilizing these methods, the project successfully projected temperature trends identified seasonality, and recognized extreme temperature events relevant to the region. The second component of the project focused on forecasting precipitation in Bloemfontein. This was a critical facet as rainfall patterns profoundly impact agricultural outcomes. To accomplish this, the study employed ARIMA and LSTM. ARIMA, a conventional yet powerful tool, was used to capture temporal variations in precipitation. LSTM, a specialized deep learning architecture for time series data, added a layer of sophistication by addressing the temporal dependencies and complexities inherent in precipitation data. The study's ability to forecast rainfall, a challenging task due to its inherent variability, holds significant implications for agricultural planning and resilience.

In tandem, the study delved into historical wheat yield data. This examination thoroughly investigates the variability of wheat yield over time in direct response to climate conditions. It goes beyond simple observation, seeking to uncover the intricate relationships between climate fluctuations and agricultural output. Wheat production is intricately linked to climate variables, and understanding this relationship is vital for agricultural sustainability. The project utilized VAR (Vector Autoregression) and Multiple Linear Regression to model the interplay between climate variables and wheat yield. VAR facilitated examining dynamic interactions between multiple time series variables, offering insights into how climate factors influence wheat production. On the other hand, Multiple Linear Regression provided a clear and interpretable model of the linear relationships between climate parameters and wheat yield. These techniques enabled the project to predict wheat yield and shed light on the nuanced interactions that underlie this vital aspect of agriculture.

The study employed an arsenal of time series analysis techniques to bridge the gap between climate variables and wheat yield. From traditional methods like ARIMA and SARIMA to the more modern Facebook Prophet, these models meticulously capture the temporal dynamics of climate and agricultural data. They provide historical insights and a foundation for forecasting future interactions between climate and wheat yield. In conclusion, integrating time series analysis techniques, machine learning models, and statistical approaches has highlighted the dynamic relationship between climate variability and wheat yield production in Bloemfontein Wheat Farms. By quantifying these relationships, we hope to support sustainable agricultural practices and contribute to a resilient food production system in the face of changing climatic conditions. This research is a cornerstone for enhancing climate resilience and fostering sustainable agriculture in the face of evolving climatic conditions. As Bloemfontein's agricultural community looks to the future, this study stands as a beacon of knowledge and guidance, helping to navigate the challenges and opportunities climate change presents.

5. Conclusions

This study delved into the intricate relationship between climate variability and wheat yield production in Bloemfontein Wheat Farms through comprehensive time series analysis. To assess the impact of climate variability, we employed a variety of forecasting models and statistical techniques. For temperature forecasting, we utilized SARIMA and Facebook Prophet. These models demonstrated their effectiveness in capturing temperature patterns and exhibited promising performance in predicting future temperature anomalies. Our analysis underscored the importance of understanding temperature fluctuations as a key determinant of wheat yield variability. Precipitation prediction was conducted using two distinct approaches: ARIMA and LSTM. The latter, is a deep learning architecture, emerged as the superior choice, showcasing its ability to capture intricate precipitation patterns and providing valuable insights into the predictability of rainfall regimes. Lastly, we employed Multiple Linear Regression and Vector Autoregression (VAR) models for wheat yield production prediction. While Multiple Linear Regression revealed valuable relationships between wheat yield and various climate variables, VAR outperformed in capturing the dynamic interplay between climate factors and wheat yield fluctuations. The outcomes from this research constitute a significant advancement in our comprehension of climate variability's impact on wheat yield within Bloemfontein Wheat Farms. The insights gleaned through this study hold substantial potential as a valuable reservoir of knowledge for the local farming community, policymakers, and stakeholders. These insights can be instrumental in formulating adaptive strategies to confront the evolving climate landscape and ensure the resilience of food security. This study underscores the pressing need for proactive measures to ameliorate the potential adverse consequences of climate change on wheat production. Furthermore, future research initiatives may further enhance these models by incorporating additional climate variables and expanding the dataset's temporal scope, thereby fostering a comprehensive understanding of the long-term implications of climate on agriculture in this region. The study methodically achieves its objectives, culminating in a comprehensive understanding of the complex interplay between climate variables and wheat yield in the Bloemfontein region.

Author Contributions:

The authors confirm contribution to the paper as follows: study conception and design: VZM and ICO; data collection: VZM; analysis and interpretation of results: VZM and ICO; draft manuscript preparation: VZM and ICO, administration and supervision ICO. All authors re-viewed the results and approved the final version of the manuscript.

Funding:

This work is funded by the DSCI-NICIS National e-science Postgraduate Teaching and Training Platform (NEPTTP)

Acknowledgments:

The authors gratefully acknowledge the infrastructural assistance provided for this research by Sol Plaatje University.

Copyright:

References

[1]. Ajilogba Caroline F. and Walker Sue. "Modeling climate change impact on dryland wheat production for increased crop yield in the Free State, South Africa, using GCM projections and the DSSAT model". In: Frontiers in Environmental Science 11 (2023), pp. 1–18.
[2]. Bonetti Sara et al. "Climate change impacts on water sustainability of South African crop production". In: Environmental Research Letters 17.8 (2022), p. 12.
[3]. Zhang Huaquan et al. "Measuring the Effects of Climate Change on Wheat Production: Evidence from Northern China". In: International Journal of Environmental Research and Public Health 19.1 (2022), p. 16.
[4]. Pequeno Diego N L et al. "Climate impact and adaptation to heat and drought stress of regional and global wheat production". In: Environmental Research Letters 16.1 (2021), p. 18.
[5]. Tofu Daniel Assefa and Mengistu Mesfin. "Observed time series trend analysis of climate variability and smallholder adoption of new agricultural technologies in west Shewa, Ethiopia". In: Scientific African 19.3 (2023), pp. 155– 171.
[6]. Liu Weihang, Ye Tao, and Peijun Shi. "Decreasing wheat yield stability on the North China Plain:Relative contributions from climate change in mean and variability". In: International Journal of climatology 41.3 (2021), pp. 2820–2833.
[7]. Shew Aaron M. et al. "Yield reduction under climate warming varies among wheat cultivars in South Africa". In: Nature Communications (2020).
[8]. Mashaba Zinhle et al. "Evaluating Spectral Indices for Winter Wheat Health Status Monitoring in Bloemfontein using Lsat 8 data". In: South African Journal of Geomatics 5.2 (2016), pp. 1–18.
[9]. Wang Lingyue, Yang Xiaoliu, and Zhao Ruina. "Assessing impact of climate change on the wheat production in the Yellow River Basin, China". In: Ciências Agrárias 39.6 (2018), pp. 2361–2374.
[10]. Govere Simbarashe, Nyamangara Justice, and E. Z. Nyakatawa. "Beneficial effect of climate change on wheat yield and water footprints in the Middle-Manyame sub-catchment, Zimbabwe". In: Journal of Water and Climate Change 123.8 (2022), 2895—2910.

[11]. Tibesigwa Byela, Visser Martine, and Turpie Jane. "Climate change and South Africa's commercial farms: an assessment of impacts on specialised horticulture, crop, livestock and mixed farming systems". In: Environment, Development and Sustainability 19 (2017), 607—636.

[12]. Obindah Gershon and Chinua Mbajekwe. "Investigating the Nexus of Climate Change and Agricultural Production in Nigeria". In: International Journal of Energy Economics and Policy 10 (2020), pp. 1-8..

[13] Tandon Ayesha. Climate change made extreme rains in 2022 South Africa floods 'twice as likely'. 2022. URL: https://www.carbonbrief.org/climate-changemade- extreme- rains- in- 2022- south- africa- floods- twice- as- likely (visited on 04/20/2023).

[14]. M Bahari and N Z A Hamid. "Analysis and Prediction of Temperature Time Series Using Chaotic Approach". In: Earth and Environmental Science 286 (2019), pp. 393–403.

[15]. Twumasi Yaw A. et al. "Time Series Analysis on Selected Rainfall Stations Data in Louisiana Using ARIMA Approach". In: Open Journal of Statistics 11 (2021), pp. 655–672.

[16]. Xianghua Wu et al. "The Development of a Hybrid Wavelet-ARIMA-LSTM Model for Precipitation Amounts and Drought Analysis". In: Atmosphere 12 (2021), p. 74.