BINDURA UNIVERSITY OF SCIENCE EDUCTAION

FACULTY OF SCIENCE AND ENGINEERING

DEPARTMENT OF STATISTICS AND MATHEMATICS

TIME SERIES ANALYSIS OF ANNUAL RAINFALL, TEMPERATURE AND MAIZE PRODUCTION IN ZIMBABWE, (1980 to 2024).

BY

DEKESHE WILIAM

B213530B

TO BE SUBMITTED IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS OF BACHELOR OF SCIENCE HOUNOURS DEGREE IN STATISTICS AND FINANCIAL MATHEMATICS.

SUPERVISOR: MS J.C. PAGAN'A

JUNE 2025

AUTHORSHOP DECLARATION STATEMENT

Title of the Thesis: Time Series Analysis of Annual Rainfall, Temperature and Maize Production in Zimbabwe, (1980 To 2024).

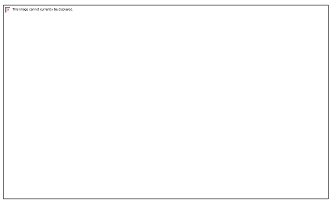
Author: **William Dekeshe**

Program: Bachelor of Science Honors Degree in Statistics and Financial Mathematics.

I, the undersigned author of the above-mentioned thesis, hereby declare that:

1. This thesis is my original work and has been prepared by me in accordance with the institution's requirements.
2. All sources, data, and references used in this thesis have been acknowledged and cited appropriately.
3. This thesis has not been submitted elsewhere for any degree or diploma.
4. I have obtained all necessary permissions for the inclusion of third-party content where applicable.

I affirm that this declaration is made with full integrity and in compliance with the institution's policies and academic practice.



Author's Signature:……………………………………………..

20-06-2025

Date: …………………….......

APRROVAL FORM

I certify, herein, that this research project is a product of my own research work and that it has not been plagiarized or copied from past sources without giving due credit. I further declare that no part of it has been submitted for another degree in this University or elsewhere.

WILLIAM DEKESHE	**………………………..**	**20-06-2025**
Student	**Signature**	**Date**

Certified by:

MS PAGANGA	**…………………..…**	**20-06-2025**
Supervisor	**Signature**	**Date**

DR.M. MAGODORA	**…….…………….**	**20-06-2025**
Chairperson	**Signature**	**Date**

DEDICATION

To my little brother Shelton.

ACKNOWLEDGEMENT

I would like to thank all the experts who shared their expertise and helped me with my dissertation. I am very much obliged to my supervisor, J.C. Pagan'a, and my coordinator, Ms. P. Hlupo, for their support and guidance. Their teaching really enhanced my work. I also thank the Bindura University of Science Education Department of Statistics and Mathematics for availing me with research facilities and administrative support. My sincere gratitude also goes to my friend, T. Chirume, whose unfaltering support, encouragement, and friendship were a constant source of strength and motivation during this work. To The Most High, I thank God for the amazing Grace, strength, wisdom, and opportunity to study. Finally, I would like to thank my family and friends for their ongoing support and encouragement throughout the process.

ABSTRACT

This study investigates the impact of climatic factors, especially rainfall variability and temperature fluctuations, on the yield of maize in Zimbabwe from 1980-2024. Using Time series analysis, the research explores the relationship between climate and maize yield, revealing a moderate positive correlation with rainfall ($r = 0.47$) and an extremely strong negative correlation with temperature ($r = -0.62$). To enhance prediction accuracy, the study compares the traditional ARIMA model with an advanced machine learning model, Long Short-Term Memory (LSTM) neural networks. Findings demonstrate that LSTM outperforms ARIMA, achieving a higher R-squared values (0.6793) and lower error metrics: 20.54% Mean Absolute Percentage Error (MAPE), 0.0584 Mean Absolute Error (MAE), 0.0057 Mean Squared Error (MSE), and 0.0752 Root Mean Squared Error (RMSE). The top-performing LSTM model forecasts a rise in maize production to approximately 2.3 million tonnes by 2025, followed by fluctuations, emphasizing the need for adaptive agricultural policy. The results serve to underscore the relevance of integrating climate data into models for improved risk management, policy-making, and increased food security in response to climate variability. Recommendations involve improving irrigation facilities and encouraging climate-resilient agriculture practices to counteract negative climatic effects and provide sustainable maize production in Zimbabwe.

TABLE OF CONTETS

List of Tables

List of Figures

List of Acronyms

1. ANN – Artificial Neural Network
2. LSTM – Long Short-Term Memory
3. ARIMA – Auto Regressive Integrated Moving Average
4. ADF – Augmented Dicky Fuller
5. Ma – Maize production
6. Ra – Rainfall Variability
7. Te – Temperature Change
8. Yr - Year
9. ACF – Auto Correlation Function
10. PACF – Partial Auto Correlation Function
11. MAE - Mean Absolute Error
12. MAPE - Mean Absolute Percentage Error
13. RMSE - Root Mean Squared Error
14. MSE – Mean Squared Error
15. AIC - Akaike Information Criteria
16. BIC - Bayesian Information Criteria
17. FAOSTAT – Food And Agriculture Organisation Statistics
18. GDP – Growth Domestic Product
19. NDS 1 – National Development Strategies 1
20. NNSNP – National Nutrition Strategy and National Policy

CHAPTER 1: INTRODUCTION

1.0 Introduction

Agriculture is among the most significant drivers of the national development, food security, creation of national income, and poverty eradication in Zimbabwe. Agriculture is connected with other drivers to provide the 2030 Agenda and The National Development Strategy 1 (NDS 1). The yield of agricultural production depending on climatic conditions has drawn private institutions, government bodies, farmers, and other stakeholders into focusing on using predictive analytics approaches to forecast agricultural trends. The trends estimates help make the right decisions, improve farm planning, and help achieve Agricultural policies under the umbrella of National Nutrition Strategy and National Policy on Drought Management frameworks.

This chapter provides the research on time series analysis of temperature, rainfall, and maize production trends in Zimbabwe from 1980 to 2024. It provides the background and rationale for the study, defines the research question and objective, problem statement in a lucid and comprehensible manner, and provides the scope, significance, limitation, assumptions, and delimitations of the study.

1.1 Background of the Study

Maize is the global most widespread grain crop. Mexico initially domesticated Maize crop in its native land about 10 000 years ago. In the passage of time between the 16th and 18th centuries maize turned out to be a well-known gain crop globally (Morris. 2004 cited in Chumo, 2013). United States of America, China, and Brazil have emerged as the world's finest and foremost maize grain producers globally. More than 1.2 billion Africans depend on maize as their staple food. Additionally, the expansion of maize production in Africa has also seen in Nigeria and then in Ethiopia, rising to become the largest producer of maize grain in the Sub-Saharan region (Eticha, 2020). In Southern Africa, Zimbabwe, among others, has also been included in the maize producing countries in the African Continent. In Zimbabwe the Comprehensive Agricultural Policy Framework unveiled (2012) unveiled that in the economy, agriculture is equally of vital importance contributes approximately 15 to18% Gross Domestic Product (GDP).

Climate change is a worldwide issue which has been impacting agricultural production globally FAO, 2020. Climate change effects on agriculture are extensive and international with temperature, precipitation, and other climatic elements influencing crop yield, quality, and availability. Climate change has brought about adverse effects on global food security with variation in rainfall impacting maize yields globally (IPCC, 2013). Maize production has declined by 10% globally due to climate (FAO, 2020). Maize production declined by 15% in Africa since the year 2000 due to rainfall shift and drought (AfDB, 2017). Climate change impacts agriculture within the region, where 75% of the crops rely on rain-fed agriculture (UNEP, 2019). Southern Africa, where Zimbabwe is located, has experienced an increasing incidence and intensity of droughts, floods and increased temperatures, impacting the production of maize (SADC, 2019). Maize yields in Southern Africa have decreased by 20% since the year 2000 due to climate factors (CSIR, 2018).

Figure 1.1: Regional maize production.

Figure 1.1 indicates that from 1985 to 2014, maize yield in Zimbabwe was typically less than 1.5million metric tonnes per hectare, while other countries such as South Africa had high yields, with over 5million metric tonnes per hectare at the period's end. Zambia and Mozambique had yields higher than Zimbabwe but lower than South Africa. Because Zimbabwe's yields are much lower, they bring down the average maize yield for Southern Africa. This means that even if some countries perform well, the regional average can still be reduced by poor performance in one country. As a result, food production in Southern Africa could become less stable, especially if more countries experience challenges like Zimbabwe.

Maize is a strategic crop and staple food crop in Zimbabwe. Maize contributes over 50% of the calorie content of the country's diet for approximately 13.1 million citizens (retrieved from USAID (2022). Maize is also used in the animal feed industry. Maize is a staple food in Zimbabwe that is primarily produced for sadza for human consumption by the citizens and for making other by-products. Zimbabwe requires about 1 800 000 tonnes per year, of which 500 000 tonnes is in hand.

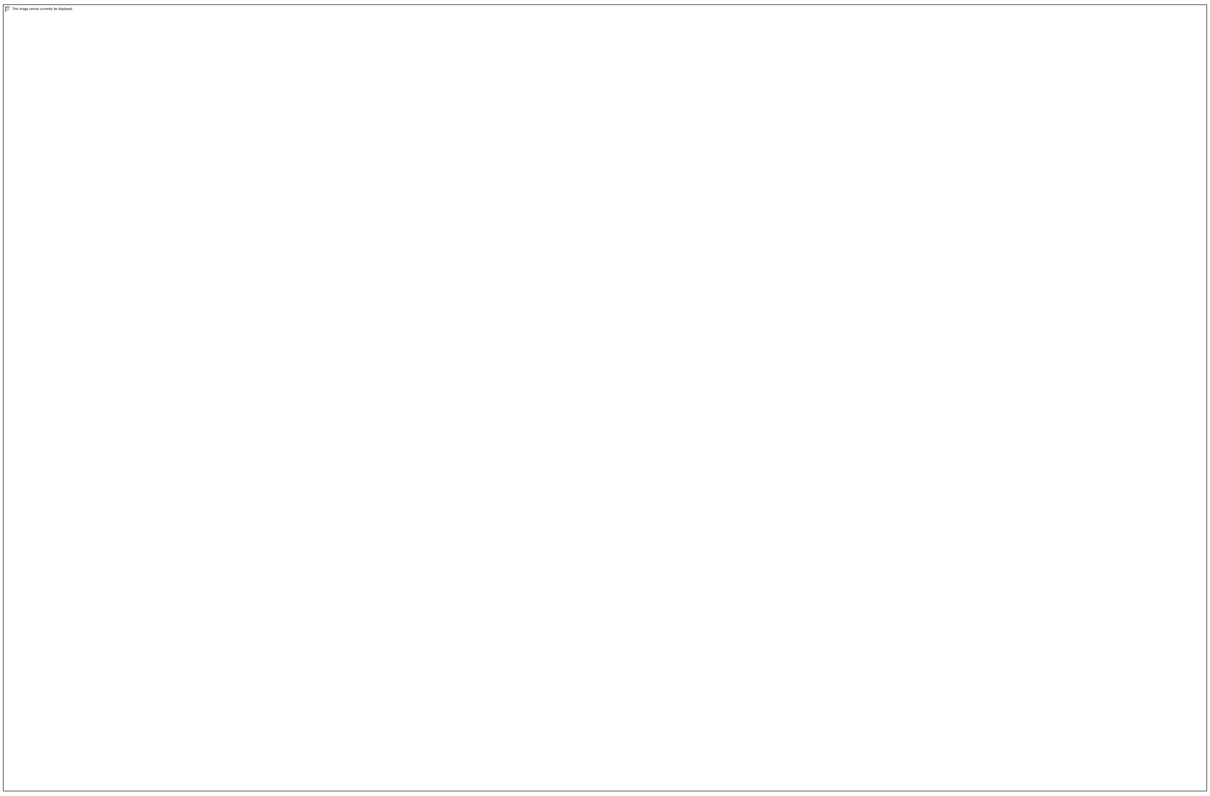


Figure 1.2: Contribution of agriculture to GDP

The pie chart shows how different farming activities contribute to Zimbabwe’s agricultural GDP. Agriculture is a major part of the country’s economy, and the chart highlights that tobacco brings in the largest share at 25%, followed closely by livestock at 24%. Maize production, which is Zimbabwe’s main food crop, accounts for 14% of the agricultural GDP, showing its important role in feeding the nation and supporting farmers’ incomes. The rest are given by cotton (12%), beef and fish (10%), sugar and horticulture (7%), and lesser values by subsistence crops and other activities. This study reveals that though tobacco and livestock predominate in terms of value, maize remains a highly important crop for economic and food security in Zimbabwe. (NAPF, 2018)

Maize is grown under rainfed during summer. Maize growing in Zimbabwe has altered since 2000 because of temperature and rainfall fluctuations. The maize yield has reduced by 30%

since 2000 (ZimStat, 2020). Maize growing was affected by floods and droughts. The year with the least recorded rainfall was 2019.

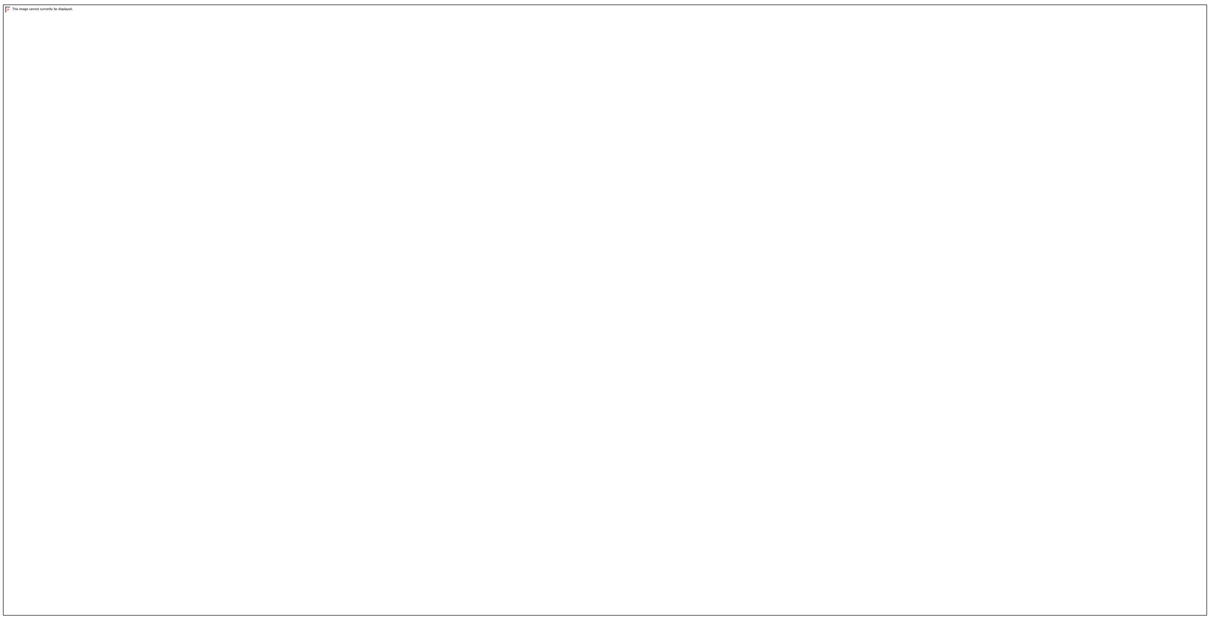


Figure 1.3: Maize production and climatic factors in Zimbabwe (1980 to 2024).

As Figure 1.3 indicates, Zimbabwe maize production is dominated by variability in rainfall. The early 1980s and late 1990s years record high levels of rainfall, following which there are spikes in production of maize, while drought years like 1992, 2015–2016, and 2019–2020 experience strong declines in production. For instance, the harsh drought experienced in 2015–2016 resulted in a major drop in maize production, as also reported in reducing by a total of about 50% for the season (WFP, 2016). Another major drop in rains was preceded by that in maize production for the period of the 2019–2020 drought (Ministry of Agriculture, 2020).

Apart from rainfall variability, Figure 1.3 shows a rising trend in temperature from year to year, as indicated by the upwards curve of the temperature graph. Rising temperatures- combined with reduced rainfall-strengthen exacerbate deteriorating maize production losses, in concurrence with previous research conducted, that rainfall variability and temperature anomalies rank among the most factors influencing maize production and food security in Zimbabwe (FAO, 2018; Chikwati et al., 2021). This requires the use of reliable time series forecasting techniques that can forecast maize production based on compound climatic factors, to inform climate adjustment planning among farmers and policymakers in risk management and food security (WFP, 2016; FAO, 2018; Ministry of Agriculture, 2020; Chikwati et al., 2021).

1.2 Statement of the problem

Zimbabwe has experienced abnormal rain patterns, rising temperatures, and prolonged dry spells. The alterations have made farming complex for the farming population to make decisions regarding planting and harvesting periods, which have led to decreased maize yields and enhanced food deficits. It has become difficult for the nation to ensure reliable supplies of adequate maize, leading to unplanned maize imports to counteract the effects on national food security. The non-linear interaction between these climatic factors has caused unclear understanding within farmers and other stakeholders on their relationship with maize output and on which best time series model to use for forecasting in order to plan ahead. This study addresses the use of advanced time series methods that provide better prediction and a clear understanding climate factors and maize production. The goal is to support smarter decision-making for farmers, policymakers, and other stakeholders, help to improve food security, and strengthen Zimbabwe's' ability to cope with clime change.

1.3 Research Objectives

1. To analyse and explore the relationship between rainfall variability, temperature change and maize production in Zimbabwe from 1980 to 2024.
2. To develop and evaluate the performance of ARIMA and Artificial Neural Networks models in predicting maize yield based on historical rainfall and temperature.
3. To compare the predictive accuracy of ARIMA and Artificial Neural Networks models in forecasting maize production.
4. To forecast maize production using Arima and Artificial Neural Networks.

1.4 Research Questions

1. What is the correlation between rainfall patterns, temperature change and maize production in Zimbabwe from 1980 to 2024?
2. How accurately can ARIMA and Artificial Neural Networks models predict maize production using past rainfall and temperature patterns?
3. Which model, ARIMA or Artificial Neural Networks, provides strength in forecasting maize production?
4. What are the estimated maize production trends in Zimbabwe based on forecast from ARIMA and Artificial Neural Network?

1.5 Significance of the study

1. To The Students

How the weather patterns alter for example on how much rain falls and the temperature which rises, gives a clear signal to the students on their impact on maize production in Zimbabwe. Learning to use forecasting tools like ARIMA and ANN, the students acquire real skills in data analysis and appreciate agricultural effects. Such skills are worth acquiring by anyone who is studying, Statistics, Economics, Agriculture or the Environment.

2. To Other Students and Researchers

Guidance on how to apply time series models to research and forecast the production of crops can be a model that other researchers or students can use or enhance for comparable studies in Zimbabwe. Even more, other researchers and students can use this research as an illustrative example for their own research.

3. To The Government

For the government, the research offers actionable information that can help in Food security planning and decision making. If only it can better predict how much maize yields, the government can plan food supplies, make use of resources more sparingly, and come up with policies to help farmers adapt to changing weather patterns. This can help immensely in preventing food shortages and allow Zimbabwe to adapt to the effects of climate change

4. To Farmers and Agricultural Stakeholders

The findings to identify the ideal planting and harvesting times for maize can be of benefit to farmers and everyone who deals in farming. Accurate forecasting, can allow them to prepare well in advance, maximize their use of resources, and avoid losses due to unforeseen weather patterns. This can lead to improved yields, generate more consistent revenues, and provide for better sufficient food for communities.

5. To Agricultural Companies and Industries

Companies like Seedco., Pioneer which trade in seeds and Sable, ZFC trading in fertilizers, and food processing companies will also benefit. With increased knowledge of future maize production, they will be in a position to foresee demand, forecast what to produce, and be able to control their supply chains. This helps them to reduce risks on markets and make improved business decisions and options.

6. To Policy Makers and Development Partners

NGOs or international institutions like FAO are policymakers who could benefit from this research. They can use good evidence when they make more programs and better policies. More efficient help can be developed through precise maize yield predictions. More importantly, they can channel their help where it is needed most, and work towards making agriculture in Zimbabwe more sustainable and resilient.

1.6 Assumptions of the study

ARIMA Model Assumptions

For this research, it was assumed that the time series data for maize production, rainfall, and temperature could either be stationary or made stationary by applying differencing techniques. This was important to ensure the reliability of the ARIMA model's forecasts. We also needed to check if the model residuals (errors) were normally distributed, had constant variance over time, and were not serially correlated with each other. Where the raw data did not meet these criteria, appropriate adjustments, such as differencing or transformation, were made in the analysis to satisfy these assumptions.

ANN Model Assumption

In using Artificial Neural Network models, it was assumed that the available historical data was enough for the model to learn and recognize the patterns linking climate variables to maize production. Data preparation steps, such as scaling and selecting relevant features, were carried out to help the model effectively learn from the data.

Other Assumptions

The study assumes the selected data is representative of Zimbabwe's maize production.

The study assumes the chosen models are suitable for predicting maize production

1.7 Limitations of the Study

Among the challenges faced during this study was the lack of local literature that specifically tackles time series analysis of Zimbabwean maize yields in relation to rainfall and temperature fluctuations. As a result, the literature review employed international research journals and articles to provide context for and support to the study. Additionally, secondary data were used in the study from reliable international sources such as FAOSTATS and the World Bank because local institutions like ZIMSTATS and AGRITEX did not have reliable complete and available historical records of maize production, rainfall, and temperature spanning 1980 to

2024. The second limitation was inability to capture all possible important variables such as farm labour, input costs, and access to technology owing to unavailability of reliable series time data for these variables. Socio-economic considerations, market forces, and policy changes were also not dealt with explicitly, which would have an effect on maize production trends but were outside the scope of the current research.

1.8 Scope (Delimitation of the Study)

To maintain a clear focus, this research concentrated on the time series analysis of maize production in Zimbabwe, specifically examining the relationship with rainfall variability and temperature changes using data from 1980 to 2024. The study also provided short-term forecasts for the years following 2024. By setting this defined time frame and focusing on national-level data, the study allowed for an in-depth analysis of recent trends and patterns, while deliberately excluding long-term projections or highly localized (provincial or district-level) analysis. This approach ensured that the findings would be relevant for short- to medium-term planning and decision-making within Zimbabwe's agricultural sector

1.9 Defined key Terms.

Agricultural activities' climate change means alterations or deviations of the mean, minimum, or maximum temperature within a time period in each region where animals or plants are raised (Harrison, L. et al 2019)

Long Short-Term Memory (LSTM) is a recurrent neural network (RNN) model that can learn from sequential data and forecast sequential data by properly capturing long-term dependencies (Greff, K. et al, 2017).

Rainfall Variability refers to the change in rain conditions, including frequency, intensity, and duration, impacting agricultural productivity (IPCC, 2013).

Maize Production Maize production involves the cultivation and harvesting of maize, a cereal crop that is vastly cultivated in all parts of the world, including Zimbabwe (FAO, 2017).

Time Series Modelling refers to the statistical technique used to analyse and forecast data arranged in time, such as rainfall and temperature information, to capture trends and patterns (Hyndman & Athanasopoulos, 2018).

Artificial Neural Networks (ANNs) are computational models that draw inspiration from the brain and are used for forecasting complex relationships between variables, such as maize yield and rainfall patterns (Gwimbi et al., 2020).

Climate-Resilient Agriculture: Farming systems and practices that are able to withstand and adapt to climate change and variability and also deliver sustainable livelihoods and food production (FAO, 2020)

ARIMA (Autoregressive Integrated Moving Average): Time series model approach used for forecasting future values from past trends analysis (Hyndman & Athanasopoulos, 2014).

Predictive Power is the ability of a model to forecast the accurate future outputs, i.e., yields of maize, based on past observations and climate variables (Shmueli, 2016).

1.10 Structure of the study

Five chapters made this research organized. 1st chapter (chapter 1) introduces the study, offering an overview of time series analysis as it applies to maize production in Zimbabwe, in relation to changes in rainfall and temperature. 2nd chapter (chapter 2) presents a review of relevant literature, highlighting previous research and the different methods used to estimate and predict maize and other crop yields. 3rd chapter (chapter 3), the research methodology is explained, detailing the data sources, variables, and model analytics specifically ARIMA and Artificial Neural Networks (ANN) used. 4th chapter (chapter 4) presents the results, performance comparisons of the ARIMA and ANN models, and provides maize production forecasts based on the model that performs best. Final chapter, Chapter 5 summarizes the study, discusses key findings and conclusions, provides recommendations, and suggests areas for future research based on the results attained.

1.11 Summary Chapter

This chapter has outlined time series analysis of Zimbabwean maize production, encompassing the background to the study, problem statement, objectives, research questions, assumptions, limitations, scope, and the research structure. The next chapter will critique literature on maize production forecasting and impacts of climatic factors, laying ground for this study's own analysis.

CHAPTER 2: LITERATURE REVIEW

2.0 Introduction

This chapter provides an overview of the relationship between climatic factors, i.e., rain and temperature, and maize yield in Zimbabwe. The chapter critically reviews literature to indicate how the variability of these climatic factors affects maize yield, an aspect of significant relevance to food security in Zimbabwe. The chapter further describes methods, e.g., time series analysis, used to measure these trends. It has declared its limitations in conducting research, principally the use of new models in climate forecasting of maize trends. Finally, it provides a framework for such research to improve understanding of how climate influences Zimbabwean maize production.

2.1 Theoretical review

2.1.1 Climate Change Impact Theory

Theory of climate change effect deals with the effect of long-term weather condition changes such as rain and temperature on agriculture. With shifting patterns in the climate, crop yields, seasons for planting, and susceptibility to disease and insects can also change (Lobell & Gourdji, 2012). For Zimbabwe, it means maize crops vary depending on the fluctuation in the climate. The theory is important in that it enables policymakers and farmers to be aware of these risks. It also educates them on how to look for alternative methods of adapting to changes, e.g., choosing varied planting dates or crop varieties. The theory enables one to have a concrete link between climate change and maize production change with the passage of time.

2.1.2 Time Series Analysis Theory

Time series analysis is a way to study data that is collected over time. It helps describe trends, patterns, and cycles in the data (Box & Jenkins, 1976). By considering the rainfall and the yield of maize year after year, the method is helpful in agriculture. Through the examination of previous data, we are able to forecast the future. One common model is the ARIMA model for time series modelling. ARIMA employs historical values, such as previous rainfall or maize yields, to assist in predicting what comes next (Hyndman & Athanasopoulos, 2018). LSTM is another new model that can pick up more advanced patterns within the data. Time series analysis is commonly used by researchers to make predictions of crop yields. Farmers and policymakers can use it to prepare for good and bad harvests. It allows interpreting how climatic factors or other factors may change maize production in Zimbabwe (Box et al., 2015).

2.1.3 Predictive Analytics Theory

Predictive analytics refers to using past data in forecasting events that may happen in the future. They use computer models for looking at what has happened in the past to predict what could happen in future periods. Brooks and Thompson (2017) explained that the first step in beginning is to find the main problem with collecting the right data and deciding on what to predict. The best information is picked to use and choose a model to make the predictions. Maize farming in Zimbabwe uses very useful predictive analytics. By looking at past climate and yield data, we can predict how much maize will be produced in the future years. This aids farmers and leaders prepare for good or bad harvests ahead of time. It also allows them to make better decisions, like when to plant or what resources to use. Climate change reduces the predictability of weather patterns and agricultural outcomes, where employing intelligent models such as ARIMA or neural networks becomes all the more (Kumar et al., 2018; Idrees, 2019). Intelligent models can manage intricate and dynamic data and improve risk management and food security.

2.1.4 Cobb-Douglas Production Function

Cobb-Douglas production function is an economic theory which illustrates how inputs interact to produce output. It was first developed by Cobb and Douglas in 1928. The model looks at inputs like land, labour, and capital. In agriculture, it can also include climate factors like rainfall and temperature. This theory helps us understand how each input affects maize yields. For example, in Zimbabwe, maize production depends not just on land and labour, but also on how much rain falls and how hot it gets. Studies have shown that changes in rainfall and temperature can make a big difference to maize output (Mupangwa et al., 2017). Using this theory, we can measure the impact of each input. This assist farmers and policymakers decide how to use resources and adapt to climate changes. The Cobb-Douglas model is a useful tool for making better decisions in maize farming in Zimbabwe.

2.1.5 Risk and Uncertainty Theory in Agriculture

Risk and uncertainty theory looks at how the farmer and decision-maker acts when the future is unclear. In agriculture, it can result from unpredictable weather, market changes, and pests (Hardaker et al., 2015). For maize farmers in Zimbabwe, rainfall is particularly uncertain: the rainy season can start late, come early, or be shorter than usual. Due to this, most farmers delay planting or opt to plant more than a single crop. These are decisions that enable them to reduce the risk of losing their harvest. Research indicates that improved weather forecasts can enable farmers to make sound decisions and minimize losses (Moyo et al., 2019). This theory accounts for the reason why accurate climate predictions and risk management measures are essential in the maize farming industry in Zimbabwe.

2.1.6 Schultz's Theory of Agricultural Transformation

Schultz's theory assert that farming gets better with new technology and improved practice (Schultz, 1964). He asserted that increased crop yield is due to accepting new ideas, employing resources effectively, and improving agriculture using modern techniques. This theory suits Zimbabwe well. Farmers are beginning to employ drought-tolerant maize seeds and improved weather forecasting equipment. These changes help them cope with climate change and keep maize yields steady. Using advanced models like ARIMA and LSTM to predict maize production also shows the impact of innovation. Schultz’s theory supports using these new tools and methods to make Zimbabwean agriculture stronger and more productive.

2.1.7 ARIMA Models

The Autoregressive Integrated Moving Average (ARIMA) model is a popular tool for time series forecasting. It uses past values to predict future values and has widespread applications in agriculture. Scientists have used ARIMA for crop yield analysis and prediction, including maize crops, using past rainfall and temperature levels. For example, Sankar and Pushpa (2019) used ARIMA to predict sugarcane yields, while Mithiya et al. (2019) used ARIMA in oilseed production. In Zimbabwe, ARIMA would be useful in interpreting how rainfall and temperature changes influence maize output over time. The model accommodates data with trends and seasonality, which is typical of weather and crop data (Box & Jenkins, 1976). ARIMA might not be effective with very complex or non-linear data, though, which restricts its application in certain situations (Mensi et al., 2017). Despite this, ARIMA remains a good starting point with regard to forecasting Zimbabwean maize production.

2.1.8 Artificial Neural Networks (ANN)

Artificial Neural Networks refer to soft replicas of the human brain. They are applied to identify patterns in data where relationships are complicated or non-linear. ANN models increasingly used for crop yields, especially crop varieties where factors such as climate, which include rainfall and temperature, interact in a highly complex manner. Studies have shown that ANN can provide more accurate forecasts than traditional models, especially when data is large and non-linear (Rather et al., 2015; Mensi et al., 2017). ANN models have been used to predict maize yields using past weather data, helping farmers and policymakers plan better (Safa et al., 2015). In Zimbabwe, where climate changes are unpredictable, ANN can help capture hidden patterns and improve forecasting of maize production.

2.1.9 Comparative Framework

Time series analysis uses Artificial neural networks and ARIMA models in forecasting but differs in their approach. ARIMA models are suitable for capturing linear relationships in data, making them effective for short-term forecasting based on historical rainfall and crop yield data (Hyndman and Athanasopoulos, 2018). However, ARIMA struggles with non-linearities and extreme weather events, which are increasingly common in Zimbabwe due to climate change (Mutasa and Nyambara, 2020). On the other hand, ANNs are better equipped to handle non-linear relationships and interactions between multiple climatic variables, such as temperature, rainfall, and soil moisture (Breiman, 2001). ANNs can adapt to new data, improving their accuracy in real-time forecasting. Studies comparing ARIMA and ANN in forecasting maize yields under varying rainfall conditions have shown that ANNs outperform ARIMA in terms of predictive accuracy, particularly during extreme weather events (Mugabe, Matarira and Unganai, 2021). The Akaike information criterion (AIC), the Bayesian information criterion (BIC), the root mean-square error (RMSE), and the mean absolute error (MAE) are some of the evaluation or performance metrics used in time series analysis to assess the performance of ARIMA and ANN models (Wheelwright and Hyndman, 2016). Mean Absolute Percentage Error, mean squared error and coefficient of determination are metrics used to compare the performances of ANN and ARIMA.

2.2 Empirical Literature Review

2.2.1 Climatic factors and Maize Production.

As the country's principal crop, maize depends heavily on adequate and timely rainfall when it is being produced (from November to March). The Zimbabwe Meteorological Services Department (2023) asserts that the country has experienced a decreasing annual rainfall since the early 1980s, with progressively deeper droughts, especially in the south. Rainfall variability is one of the most intrinsic climatically driven factors affecting agricultural output in Zimbabwe, particularly for maize, which is highly sensitive to rainfall variability (Matarira, Unganai and Mukarakate, 2019). Matarira, Unganai and Mukarakate (2019) confirmed that rainfall variability was responsible for the majority of maize yield decline over the last decades. Their study, using 1980-2018 historical rainfall records, revealed that erratic rainfall was responsible for variation in maize yields. Chagutah (2021) also affirmed that repeated droughts, particularly during the El Niño season, were largely to blame for reducing maize yields. Both studies reveal that unpredictability stems from rainfall variability. Research indicates that irregular rainfall patterns, both droughts and instances of excessive rainfall, have a significant impact on maize production(Abaye, et al. 2019).

Rurinda et al. (2014) noted that maize yields in the semi-arid areas of Zimbabwe go through declines during drought years, with more than 40% of crop losses in extreme cases. The authors also demonstrates that climate change has amplified these fluctuations in rainfall, rendering maize production more uncertain. As per Mutasa and Nyambara (2020), there could be a possibility of crop loss in the event of irregular rain patterns, especially during the peak cropping periods. Mushore et al, (), studied the relationship between meteorological normal and maize production.

The study read through that temperature is increasing while rainfall is reducing over time thereby raising chances of low maize yield in mount Darwin Zimbabwe. The researcher also observed significant negative correlation of R= -0.905 between maize yield and number of dry spells and high positive R= 0.777 between length of season. The researcher also observed a strong positive correlation of r=0.753 between percentage normal rainfall and percentage normal maize yield Climate research in Zimbabwe consistently depicts warming temperature effects on maize production which are usually afterward compounded by coincidental deficit of rainfall.

Moyo, Mvumi and Kunzekweguta (2022) also evaluated climate trends and maize productivity in selected agro-ecological zones in Zimbabwe. They noted that raised mean seasonal temperature, especially at decisive growth phases, had a negative correlation with maize productivity. Moreover, effects were more severe in those seasons witnessing less-than-average rainfall. This indicates the interactive stress of dryness and heat, a key consideration for your forecasting models that incorporate rainfall and temperature.

Similarly, research conducted by Chikwati, Nciizah and Bangira (2021) of semi-arid areas of smallholder farming systems in Zimbabwe documented that occurrences of heat stress, even with adequate cumulative rainfall, could lead to yield penalties due to impacts on pollination and grain filling. Mapfumo et al. (2023) conducted a research study of the impact of climate variability on maize and other crop suitability in Zimbabwe. Their findings, employing climatic projections, suggested that rising temperatures were likely to reduce the amount of land favourably conducive to peak maize production, even if rainfall patterns stayed relatively constant in some regions, highlighting temperature as a single and future constraint factor.

2.2.2 Time Series Analysis in Agricultural Research

Time series analysis is a statistical method widely used in agricultural studies to analyse long-term data and forecast future trends.

In Zimbabwe, time series models have been applied to investigate the relationship between rainfall variability and maize yields. Mapuwei, et al., (2022) carried out a study using a time series ARIMA forecasting model in Zimbabwe for tobacco production. A time series of Zimbabwe tobacco yield was employed in order to design appropriate methods to tackle this declining pattern. In order to forecast the tobacco yield for 2019 to 2023, they focused on the yield from 1980 to 2018 to build ARIMA models. Tobacco yield showed a declining trend that changed slightly over the projected years. In order to improve tobacco production researchers recommended the adequate provision of inputs, farmers education and assistance from the government regarding tobacco production. They also recommended other academics to use alternative techniques such recently developed ANN models for comparisons and coming up with the best model. Gwimbi (2009) analysed climate change and its impact on cotton production in the Gokwe district of Zimbabwe, using a time series analysis of temperature and rainfall for a period of 30 years.

Correlation tests between the independent variable climate and the dependent variable cotton output were assessed in order to determine the nature and strength of the relationship. The researcher discovered that cotton production levels decreased with the reduction in precipitation and increase in temperature in the district. The same results further showed that farmers are highly vulnerable to climate change. The study reviewed the need to invest in climate adaptation strategies to help them cope better and reduce their harm to climate change. Mutasa and Nyambara (2020) applied ARIMA models to forecast maize production in Zimbabwe, using historical rainfall data from 1980 to 2018. While the model was effective in predicting short-term trends in maize yields, it struggled to capture the non-linear relationships that occur during extreme weather events, such as droughts or floods. This limitation points to the need for more advanced models such as neural networks, that can handle the non-linearity of such data.

However, it has been pointed out by Hyndman and Athanasopoulos (2018) that ARIMA models are weak in describing non-linear relationships especially in cases of climate variability and crop production. This weakness in the tool has compelled researchers to dive deeper into other advanced predictive models that could be applied for agricultural forecasting.

2.2.3 Advanced Predictive Models for Maize Yield Forecasting

Mugabe et al. (2021) used ANNs to develop a maize yield forecasting model for Zimbabwe.

The researchers developed artificial neural network models to predict maize yield based on various climatic variables, demonstrating the effectiveness of ANN in capturing complex relationships between these variables and maize productivity. Key climatic factors such as temperature, rainfall, and humidity were identified as predictors of maize yield. The study emphasized the importance of these variables in influencing agricultural outcomes in Zimbabwe. The ANN models showed high accuracy in predicting maize yields, outperforming traditional statistical methods Arima. This suggests that machine learning techniques can provide more reliable forecasts for agricultural planning and management. The authors recommend further studies to incorporate additional variables and refine the models, which could improve prediction accuracy and applicability across different regions and climatic conditions in Zimbabwe. A study done by Ghodsi, et al., (2012), used artificial neural networks to estimate Iran's wheat production. ANN model used as a prediction tool, and they selected eight important wheat producing elements to carry out their review since the negative effects of soil and climate on wheat output in many areas of Iran.

Annual data span from 1988 to 2006 used in their study. The authors used the model hundred times to minimize noise. The forecast of five years period was made on wheat production, and the results were such that forecasting can be very useful for informed decision making and supply chain optimization, thus recommending the use of ANN model as good in use for the forecasting of wheat production. The study recommends further research on machine learning in other predictions Munyati and Jirah 2020) examined the use of machine learning algorithms in predicting maize yields in Zimbabwe. The study used historical climate data and satellite imagery to train several machine-learning models, such as random forest, support vector machine, and neural networks. The results of their study showed that the random forest model performed better in predicting maize yields, with an accuracy of 87.5%.

2.2.4 Climate Change and Future Projections

Kożuch, et al., (2023), did a comparison of ANN and traditional time series models for timber price forecasting and analyse the average price decrease for timber in Poland. The research used quarterly time series of timber prices from 2005–2021.

ANN was found to be far superior to ARIMA and Prophet Model in forecasting price fluctuations. The research suggested that precise price prediction is crucial in order to comprehend the market variables with respect to suppliers and customers. The researchers motivated future research to keep on considering ANN -models for formulating the optimal solution supporting accurate forecasting for problem-solving. The above projections indicate the role of predictive modelling in the formulation of strategies to mitigate the effects of future climatic shifts on maize production.

2.3 Research gap

Past studies on agricultural production in Zimbabwe have important limitations. Many studies only looked at the short-term effects of either rainfall or temperature, not both factors (Mugabe et al,2022). These studies do not clearly give an understanding on how these two climate factors interact with maize production, especially when climate change becomes more unpredictable. The use of ARIMA models has been common, but these models do not work well when the data is complex or non-linear and often occurs with weather and crop data(Hyndman and Athanasopoulos (2018),). Some recent studies have used Artificial Neural Networks (ANN) and other machine learning models. However, very few studies have directly compared ANN and ARIMA for predicting maize yields with long-term data. Little toolkit also exists on which model, ARIMA or ANN, is optimal for forecast agricultural production in Zimbabwe. Farmers and policymakers require unambiguous guidance on what models to rely on for planning and food security.

2.4 Proposed Conceptual Framework

The relationship between maize production, its restrictions, and impacts are proposed by the conceptual framework below.

Figure 2.1 Proposed Conceptual Framework.

Temperature, soil type, agricultural practises, seed quality, government support, labour and macroeconomic factors affects maize production. Improvements in maize production will results in several constructive outcomes, including increased foreign exchange revenue, maize self-sufficiency, reduced poverty, increased employment, and an improved government budget.

2.5 Chapter Conclusion

This chapter has provided an important review of existing literature and demonstrated its importance to the current study. The research shows that rainfall variability and temperature changes have a major impact on maize production in Zimbabwe. Both time series models, ARIMA, and machine learning methods, like Artificial Neural Networks, have been widely used to predict maize yields using past climate data. ARIMA models are useful for understanding short-term trends, while machine learning techniques are better at capturing complex patterns to come up with accurate predictions, when data is non-linear. The next chapter will outline the research design, including data collection and the methods to be used for analysis.

CHAPTER 3: METHODOLOGY

3.0 Introduction

Research is the process of fact-finding that entails a scientific study or comprehensive research on a given topic. Research has been shown to be an essential tool that forms the foundation of government institution and policy-maker's economic decision making (Mackey & M 2013)

This chapter clearly outlined the data sources, population targeted and sample, methods of data collection, description of variables and their expected relationships, data analysis procedures, as well as the process of model development and selection. Research methodology ensure that the study is conducted in a structured, reliable, and valid manner.

3.1 Research Paradigm

A dual adoption of research paradigm was done in this study. This was to effectively analyse the influence of climatic factors on maize in Zimbabwe. The time series analysis is based on the positivist paradigm because it enables objective measurement of the relationship between rainfall patterns and maize yields by way of making predictions using quantitative techniques (Saunders et al., 2019). This approach ensures empirical reliability, and developments for time series models (ARIMA model) can be created, establishing repetitive temporal trends and patterns. The aim is to objectively analyse data and test hypotheses related to the impact of climatic factors on agricultural output, aligning with the core tenets of positivism (Bryman, 2023).

Furthermore, the constructivist paradigm was incorporated through the application of artificial neural networks, (LSTM model). ANN are well-suited to capture the complex and nonlinear dynamics in the data. By integrating historical maize production data with rainfall variability, these models adaptively learnt and enhanced the predictive accuracy of future yields.

3.2 Research Design

Research design is a work plan detailing how the research was to be undertaken, type of data to be collected, tools and techniques that were to be employed to obtain data and the method of data analysis to be used (Wyk, 2012). This design is suitable for examining trends, patterns, and relationships in data collected sequentially over an extended period (Box et al., 2015). Advanced statistical techniques and an extensive sample size were employed to guarantee the precision of the results and the extensibility of the results. The design offers rich insights to inform evidence-based agricultural decision-making.

3.3 Data Collection and Sources

Hamed T., (2021) has described data collection as the procedure for gathering data with the aim of acquiring insights into the research question. The collection of data for the study was non-experimental. It was based on secondary data from concerned publications and websites. Data on annual maize production (in tonnes) for Zimbabwe from 1980 to 2024 was collected from The Food and Agriculture Organization Statistics (FAOSTATS). FAOSTATS is a specialized agency of the United Nations that focuses on reducing poverty and hunger. The statistics are available on their official website. Moreover, national average yearly rainfall trends were obtained from the World Bank Indicators. It is an open-data platform, whose statistics are freely available to any person interested in statistical analysis. The initial step was to get logged into the data portal with necessary variables and period for the research. Maize yield data and rainfall trends were located and accessed. The collected data was then copied into an Excel sheet for ready access and arrangement during the construction and validation of predictive models. The table 3.1 clearly shows the sources of data collected and period.

Table 3.1 Collected Data and Time Frame

Data Type	Source	Time Frame
Rainfall Variability (mm/year)	**World Bank Indicators**	**1980 to 2024**
Maize Production (tonnes/year)	**FAOSTATS**	**1980 to 2024**
Temperature change (°C)	**FAOSTATS**	**1980 to 2024**

Justification of data sources

Secondary data sources utilization has advantages in scholarly research, most importantly through the reduction of data collection time and expenses (Hamed T., 2021). FAOSTAT and World Bank Indicators are highly academically relevant. They are used as credible and comprehensive sources of international statistics. FAOSTATS and World Bank Indicators, give data that is current, screened, and is capable of handling various sample sizes. This makes it an excellent resource for academic research and analysis.

3.4 Population and Sample Period.

Riya, (2023) referred population as the entire group or set of individuals, objects, or events that possesses specific characteristics and are of interest to the researcher. Population: This study includes historical maize production records throughout all provinces of Zimbabwe. A sample is a representative subset of this population and is carefully selected to facilitate robust analysis, according to Riya (2023). Sample is data on maize production from the period 1975 through to 2023. Employing a purposive sampling technique, data exhibiting critical attributes, such as data points, discernible patterns, and production trends, was meticulously selected. This methodological approach ensured the suitability of the dataset to meet time series analysis conditions.

Justification

The 45year timeline (1980 to 2024) were chosen because this long period aids in show trends in climatic factors and maize production in Zimbabwe over time. Looking at many years helps us see patterns, big weather events, and changes in agricultural output. Using a lot of data also makes the predictive models perform better and gives us more reliable results (Hyndman and Athanasopoulos, 2021).

3.5 Data Validity and Reliability

Validity and reliability of data are crucial in ensuring the credibility and quality of any research study. These are the pillars of the scientific accuracy of the procedure of data analysis and enhance the plausibility of the outcomes.

Data Reliability

Data reliability refers to the robustness of data in being coherent, consistent, and replicable multiple times over by other sources (M Hughes, 2015). For this study, FAOSTATS and Word Bank Indicators maintain reliability of their data through scheduled updates and validation. This way, the data is guaranteed to be free from serious errors and biases and inconsistencies and hence ideal for modelling rainfall variability and dynamics maize production.

Data Validity

Data validity refers to the degree to which the data reflect the concepts they are designed to measure (Heale R., 2018). For FAOSTAT and World Bank Indicators, it is achieved through adherence to established standards and methodologies; the datasets also agree with other official sources. Thus, in analysing the relationship between rainfall variability and maize production, the choice of accurate data reflected reality.

3.6 Tools of Research

Research instruments are the tools or mechanisms used to collect, measure, and analyse data within a study (Bhaskar SB., 2016). A search engine Google, assisted with accessing and extracting historical rainfall data and maize production data from reputable sources, FAOSTATS and World Bank Indicators using a laptop. To ensure methodological rigor, the data underwent preprocessing that included cleaning, validation, and standardization with the intent of reducing errors and inconsistency using statistical package R-Studio and Microsoft Excel. These software packages enhanced computational accuracy, results, and conclusion of this study.

3.7 Variables Description

This study analysed three primary variables, rainfall variability, temperature change and maize production. They are all defined and measured to align with the research objectives. Figure 1 below illustrate the description of variables.

Table 3.2 Description of Variables

Variable	Symbol	Unit of Measurement	Role in Model	Source
Maize Production	Ma	Metric Tons	Dependent Variable	FAOSTAT
Rainfall	Ra	Millimetres (mm)	Independent Variable	WORLD BANK
Temperature Change	Te	Degrees Celsius (°C)	Control Variable	FAOSTAT

Justification

Rainfall variability is an independent variable that is vital since it has a direct influence on maize production and hence determining its trends and anomalies between 1980 and 2024 based on historical data is important. Maize as the most important staple crop and agricultural economies in Zimbabwe, its production, is the dependent variable. The yield patterns at the annual scale indicate a requirement for understanding the effect of rainfall because maize in Zimbabwe is predominantly rain-fed. Temperature change has been introduced as a control variable as it influences crop growth, evapotranspiration rates and the compatibility of the climate with maize in general.

3.8 Data Analysis Procedures

The data analysis followed a structured process, encompassing diagnostic tests, model building, and evaluation.

3.8.1 Data Preparation

According to Chen et al., data preprocessing, including handling missing values and data normality test, are important in improving the accuracy and efficiency of model. Missing values were checked using R function colSums(is.na(DATA)).

Normalisation

Min-max normalisation was used in this study, allowing data from different ranges to be standardised to a scale of 0 to 1. The formular for Min-max is show below. $X^{\prime} = \frac{X - X_{min}}{X_{max} - X_{\min}}$:

$X^{\prime}$ represents normalised value

X is the original value

X_{max} the maximum value of the data set

X_{min} the minimum value of the data set

Normalisation is an essential data processing procedure that standardises data to mitigate the influence of varying data distribution.

Data Partitioning

The data was segmented into two sets for the purpose of training and testing the model. The training set encompassed the biggest fraction of the sample data and was used to capture the present patterns in data. 70% of the data was used in the training set whilst the remaining 30% was used as the testing test. Sample data was fractioned as follows:

1. Training set = 80% of total sample size

2. Testing set = 20% of the total sample size

3.8.2 Diagnostic test

Prior to model development, several diagnostic tests were conducted.

Stationarity test

Time series stationarity is the concept where mean (μ) remains stationary, variance (σ^2) is constant and there is no periodic pattern or seasonality (QuantInsti, n.d). Therefore, in the case of maize yield, rain variation and temperature variation data, variance and mean must remain constant for them to be stationary. Non-stationary data can lead to spurious regressions and incorrect forecasts (Brooks, 2019). Violation of conditions of stationarity was controlled to maintain stationarity using the Augmented Dickey-Fuller Test (ADF). The ADF formula utilized is given as follows,

$$\Delta y_t = \alpha + \beta_t + \gamma y_{t-1} + \sum_{i=1}^{n} \varphi_i \Delta y_{t-i} + \varepsilon_t$$

Where:
Δ is the difference operator denoted as $\Delta y_t = y_t - y_{t-1}$

y_t represents maize production at time t

α is a constant which reflects the average level of production when other variables are held constant

β is the coefficient of time trend that captures any systematic trend in maize production over time

γ is the coefficient of the lagged term that assesses the influence of previous maize production in current levels

φ_i is the coefficient of the lagged differences

ε_t is the white noise or error term, representing unobserved factors affecting maize production.

Normality test of residuals

normality of the residuals was performed with Shapiro-Wilk test. This was done to verify that residuals are usually distributed, one of the significant assumptions of this statistical tool. p-value > 0.05 accepts normality assumption.

Independence Test

A key diagnostic instrument in time series analysis, independence test was used to analyse the independence and lack of autocorrelation in rainfall variability and maize production data residuals. Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) plots are graphical devices utilized to select any statistically lags, where a lack of any spikes within the confidence intervals demonstrates independence.

Homoscedasticity Test

To ensure that error variance remains homoscedastic on variables homoscedasticity test was utilized. Breusch-Pagan, Goldfeld-Quandt and White test are statistical tests utilized in order to assess homoscedasticity. Breusch-Pagan test was used in this research for visual inspection of errors or residuals. The formula used by Breusch-Pagan's for the test statistic is, $BP=n \cdot R^2$, where n represents the number of observations and R^2 is the value of the squared residuals. BP is followed by a chi-squared (χ^2) test with the number of independent variables as degrees of freedom. An example of $BP<\chi^2$ should exhibit homogeneity to be present. A p-value > 0.05 showed that the residuals display constant variance (homoscedasticity), assists in precise models' specification and inference.

3.8.3 Analytical Model

Analytical models are used to simulate, explain, and predict the process of the physical mechanism in complex physical processes. Autoregressive Integrated Moving Averages (ARIMA) model and Long-Short Term Memory neural network was addressed in this section to achieve the goals and research questions of this study.

Starting with Arima model, one of the most popular, classic and effective time series models. (Rathod and Mishra, 2018. It is a combination of three components clearly named and explained below,

Autoregressive (AR) models forecasted future values based on a linear combination of the previously observed values. The approach is represented by:

Where y is the time series being forecasted at multiple times, φ is the coefficients of the lags, ϵ is the error term often normally distributed and p is the number of lagged components, also known as the order.

Integrated (I), the middle part of ARIMA is the number of orders d of differencing required to make the time series stationery.

Moving Average (MA) models this is the last part, forecasted maize production depending on past forecast errors instead of the observed actual numbers. Denoted

Where y is the time series being forecasted at multiple time steps, θ are the coefficients of the lagged forecast errors, ϵ are the forecast error terms and q is the number of the lagged error components.

The ARIMA model's standard form is then written as $ARIMA(p, d, q)$ where the letters p stands for the autoregressive term order, d for the differencing term order and q for the moving average term order, Arunraj et al., 2016; Hyndman and Athanasopoulos, 2018. Mathematically, the Arima model can be expressed as (Awe et al.,2020),

Where p, d, q are the coefficients of AR, I and MA parts respectively, ϵ_t is for the error term of Y_t that is supposed to be uncorrelated. Δ is the shift operator, $\varphi_1, \varphi_2, \dots, \varphi_p$ are the AR parameters. $\theta_1, \theta_2, \dots \theta_q$ represents the MA section.

Model Identification

The first step of stationarity check should be achieved using the Augmented Dicky-Fuller technique to ensure that mean and variance for rainfall variability and maize production data is consistency. Differencing was applied to the non-stationary data in order to stabilize the mean and prepare it for analysis. Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) were employed to determine the initial values for the orders (p,d and q) of AR and MA processes. The analysis identified peaks in Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) highlighted potential lags in the data.

Parameter Estimation

Estimation of parameters involves estimation of Arima models and checking various coefficients between them. is to select a stationary and parsimonious model whose coefficients are , and its fit is good.

Diagnostic Checking

This model estimated should be subjected to a diagnostic check to ensure that time series analysis conditions are met. Stationarity of data should be achieved. Independent, Residual was achieved by examining ACF and PACF (Namin. S, 2018). Shapiro-Wilk tested for should also be normality of residuals. The Breusch-Pagan test was used for homoscedasticity test. The goodness of fit was examined using Akaike Information Criteria (AIC) and Bayesian

Information Criteria. The most performing model with the lowest AIC and BIC should be used (Enayaba et al., 2024).A skeletal structure of Arima model development shown in Figure 3.1

This image cannot currently be displayed.

Figure 3.1 Skeletal structure for Arima model development

Long Short-Term Memory Neural Network

In recent times, there have been advancements in new sophisticated predictive methods in time series (deep learning) that were created to handle the complexities involved in the forecasting models. LSTM is a particular type of Recurrent Neural Network (RNN) method that was originally proposed by Hochreiter and Schmidhuber (Hochreiter and Schmidhuber, 1979).

LSTM is a modified and special kind of Recurrent Neural Network (RNN) with the capability of remembering the values from earlier stages for the purpose of future use, (Patterson, 2017). Traditional RNN, struggle with vanishing and exploding gradient problem, which is a default behaviour for LSTM. LSTM can better describe dependencies in sequential data by efficient storing and retrieving information over extended periods of time, (Reddy et al., 2022). It has already found application in most disciplines and includes future agricultural productions and climatic variables.

LSTM Architecture.

The architecture of LSTM includes the memory cell that holds information for a long duration, which is regulated by three gates that include, input, forget and output gate (Tian et al., 2021). The gates are on sigmoidal neural network layer, which facilitate cells to optionally permit the data to flow, stored and discarded. the internal structure of LSTM (Colah's Blog, 2015) is shown below.

Figure 3.1 Structure of LSTM (Colah's Blog, 2015) extracted from https://www.geeksforgeeks.org

Input gate added relevant data and controls how much fresh information (that is maize production and rainfall variability data) need to be in the memory cell and a $tanh$ layer makes a vector of new candidate values that could be added to the state. The equations are given by:  Where the output y_t is obtained by applying an activation function sigmoid σ to the product of the weight matrix W_t and the input vector x_t, followed by the addition of a bias term b_i. This $y_t$represents the output of the input layer at time t.

Forget gate, irrelevant data was removed from the memory cell. The equation is given by: $f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f)$.

Where Two inputs x_t (input at the particular time) and h_{t-1} (previous cell output) are fed to the gate and multiplied with weight matrices followed by the addition of bias. The resultant passed through an activation function which gave a binary output. Zero output cell state means the data is forgotten and for output 1, the information is retained for future use.

Output gate decided the yield from each cell and extracted relevant data from current cell state using the output equation $h_t = o_t * \tanh(C_t)$. The activation function for output gate is $o_t =$

Where the output gate function o_t presented by applying tanh function in order to output a vector, data is regulated by the sigmoid function σ, which decides what to retain based on the previous hidden state h_{t-1} and present input x_t and including the bias term b_o. Final output obtained by multiplying regulated values with tanh vector.

Data preprocessing

Feature Engineering

Lagged variable (Lag1_Ra, Lag2_Ra, Lag1_Te, Lag2_Te) and rolling statistics (RollingMean_Ra, RollingSD_Ra, RollingMean_Te, RollingSD_Te), were created from the normalised data. This enhances LSTM model with more contextual information about past trends and variability.

Sequence Creation

The data was transformed into sequences that were best for LSTM model input. 1 was the best length to use on the sequence length. This meant that for each input sequence, X consisted of features ("Ra", "Te", "Lag1_Ra", "Lag2_Ra", "Lag1_Te", "Lag2_Te") at time $t-1$ to predict Ma at time t.

LSTM Architecture and Training

The LSTM model was built upon the Elman network type from RSNNS package in RStudio.

The backpropagation through time (BPTT) method was utilised to update the weights of the LSTM neurone weights before training the model with the training data. This involved computing the gradients of the loss function with respect to the weights using the chain rule (Sadowski, 2016). Modifying iteratively the weights based on the estimated gradients, increasing capacity for precise prediction, the LSTM model learns to improve its performance on the training data (Bhimavarapu et al., 2023).

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial y} * \frac{\partial y}{\partial h} * \frac{\partial h}{\partial W}$$

Hyperparameter

A crucial in optimising model performance. The first and second moments of the slope are considered through the application of Adam algorithm, an optimisation technique. This approach enabled the algorithm to adjust the learning rate for each weight independently, thereby improving the training process of the LSTM model (Bhimavarapu et al., 2023)

$$v_t = \beta_1 * v_{t-1} + (1 - \beta_2) * \left(\frac{\partial L}{\partial W}\right)^2$$

$$W = W - \alpha * \frac{m_t}{\sqrt{v_t} + \epsilon}$$

Where m_t and v_t are first (mean) and second (variance) moment estimates of the gradient respectively. W is the weight being updated and α is the learning rate. β_1 and β_2 are hyperparameter representatives that controls the exponential decay of the first and second moment estimates respectively, and ϵ represents a small preventer to zero division. Different hyperparameter settings were tested to come out with the optimal configuration. The best model, achieving the lowest RMSE was used.

3.8.4 Metrices Used for Model Evaluation

Statistically, parameter diversity is used in the measurement of the residuals while verifying accuracy in the model. The most important measures of performance being considered were Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Adjusted R-squared and AIC and BIC. They were measures of the predictive strength of the model and how it was suited to fit the observed data.

Mean absolute percentage error is computed as the total of individual absolute errors divided by actual values, then averaged over all observations. This indicates how far the predicted values of the hypothesized model differ from the respective actual values. It will benefit stakeholders in agriculture that will appreciate expression of error in percentage points. Mathematically, it is given as: (De Myttenaere et al., 2016):

$$MAPE = \frac{1}{m}\sum_{i==1}^{n}\frac{|y_i - \hat{y}_i|}{y_i}$$

y_i represents the exact values, $\hat{y}_i$ denotes the forecasted figures, and m is the total number of observations.

Mean absolute error (MAE) is defined as the arithmetic mean of the absolute differences between the actual values and the anticipated values. It provides a straightforward measure and gives a linear score that's easy to interpret. It provides a clear indication of the average error in the same units as the data for example, tons of maize (Hodson TO., 2022)

$$MAE = \frac{1}{m}\sum_{i==1}^{n}|y_i - \hat{y}_i|$$

Root mean square error (RMSE) is the standard deviation of the standard deviation of the residuals, providing a measure of how well the predictions are concentrated around the best fit line (Hodson TO., 2022). It gives more weight to larger errors due to the squaring of the residuals. This benefit if larger errors are more critical in the context of agricultural anticipations. Mathematically expressed as:

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

R-squared (R^2) measures explained variation. Shows the proportion of the variance in the dependent variable being predictable from the independent variables (Gao, J., 2024.)

Mean Squared Error (MSE) Measures the average of the squares of the errors. It penalises bigger errors more heavily (Hodson TO., 2022).

Akaike Information Criterion (AIC) And Bayesian Information Criterion (BIC)

Akaike information criterion and Schwartz-Bayesian information criterion estimates model quality and evaluates and assesses model quality. Both balances complexity and fit and penalizes models with high numbers of parameters in order to prevent overfitting, but BIC imposes a stronger penalty on complexity than AIC. The best preferred model yields the least AIC and BIC. The two equations below show AIC and BIC.

$$AIC = 2k - 2\ln(L)$$

$$BIC = \ln(n) - 2\ln(L)$$

Where L is the likelihood of the model, k is the number of parameters and n is the number of observations.

3.9 Ethical considerations.

Relevance of data: This study required access to relevant data. The process of collecting data took into consideration ethical considerations. Permission to use the needed data came from appropriate sources. For instance, permission was requested from FAOSTATS and World Bank Indicators by signing up and logging in using google mail.

3.10 Summary chapter

This research utilized predictive models ARIMA and LSTM to explore the impacts of rainfall fluctuation and temperature change on maize yields in Zimbabwe. Ethicality was among the key elements to the research process, promoting ethical and open data use. Requests were made with the respective authorities for access to rainfall and production data, with adherence to the set ethical standards. Data has been thoroughly pre-processed using preprocessing of missing values, normalization, checking stationarity, independence, and homoscedasticity in order to improve the performance of the models. Both ARIMA and the LSTM models have been employed because the former demonstrated its strength in using time series while the latter is a promising candidate in identifying the subtlest patterns and long-term dependencies within the data.

Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Adjusted R-squared and AIC and BIC, were conducted in model selection. These measures ascertain the most efficient predictive model. The findings give valuable feedback concerning the interaction of climatic variables with agricultural yield, giving a solid foundation for sound decision-making for the reasons of climate resilience and agricultural planning. The next chapter focuses on data presentation, analysis, discussion and chapter conclusion, they all expand on the foundation of methodology.

CHAPTER 4: PRESENTATION OF DATA, ANALYSIS AND DISCUSSIONS

4.0 Introduction

This chapter provides an integrated presentation, analysis, interpretation, and discussion of findings as a way of answering research questions and objectives. The time series analysis used in examining maize production in Zimbabwe, subject to rainfall variability, enabled the derivation of insights towards effective discussion and conclusions.

4.1 Data Overview Statistics

Table 4.1 Descriptive Statistics

Statistic	Ma	Ra	Te
Count	45	45	45
Mean	1414268	666.0695556	0.5135111
Median	1469664	680.950	0.426
StdDev	623655.1	123.6489064	0.4693158
Min	361900	447.16	-0.364
Max	2833395	941.18	1.523
Range	2471495	494.02	1.887
Skewness	0.44142	0.1029274	0.434945
Kurtosis	-0.547859	-0.5141214	-0.7439913

Descriptive statistics of rainfall (Ra), temperature (Te), and maize production (Ma) in Zimbabwe for 45 provides an accurate description of their relationship or association with variables, variability, distributional feature, and effect. Mean rainfall: 666.0696mm, standard deviation is 123.65 mm, shows variability. Temperature anomalies skewness: 0.434945- moderate positive skewness std. dev: 0.47°C shows persistent warming trends. Maize production: means around 1.41 million tons with very high standard deviation of 623,505.5 tons indicating flat action in the yields. Moderate positive skewness in rainfall as well as maize production suggests occasional years of better yields but steady trends. Platykurtic distributions of all three variables suggest fewer extreme occurrences. Range of maize production is unusually large (2,471,495), supporting the volatility in agricultural output. Variability in rainfall is also high, with a range of 494.02 mm, which might influence planting and harvesting seasons. The control variable Te ranges from 1.887, indicating that it has partial moderating effect on some of the factors of maize yield prediction.

Looking at the minimum and maximum values, maize production ranges from 361,900 metric tons to 2,833,395 metric tons, demonstrating a considerable difference across years, due to varying climatic conditions and agricultural practices. Similarly, rainfall fluctuates between 447.16 mm and 941.18 mm, emphasizing the unpredictability of Zimbabwe's climatic conditions. Te varies from -0.364 to 1.523, indicating its influence across different seasons.

4.2 Annual Trend Analysis

4.2.1 Maize annual trend

Figure 4.1 Maize annual trend.

Figure 4.1 shows that maize production in Zimbabwe went up and down sharply from 1980 to 2024. The peak was mostly achieved in the first few years of 1980s, but since the 1990s, production became more unstable and overall lower. There are many years of extreme falls, punctuated by short periods of improvement, suggesting maize output has been unstable and typically below past peak levels during the past decades.

4.2.2 Rainfall variability annual trend

Figure 4.2 Rainfall variability annual trend

Figure 4.2 indicates that Zimbabwe rainfall from 1980 to 2024 has an irregular pattern with big rises and falls from year to year. There were some years of very high rainfall and others of very low rainfall, especially around 2000 and 2017. Over the long term, there is neither a decreasing nor increasing trend, but the pattern illustrates that rainfall is very variable and not stable over the years.

4.2.3 Temperature change annual trend

Figure 4.3 Temperature change annual trend

The Zimbabwean temperature has largely increased from 1980 up to 2024, with a lot of fluctuations in between. The graph illustrates that, even though there have been a couple of low-temperature years, the years later have higher and more frequent peaks. This is a sign that temperatures are rising over the years, with more extreme hot years, particularly from around 2000.

4.2.4 Maize production, Rainfall variability and Temperature Change Combined Trends

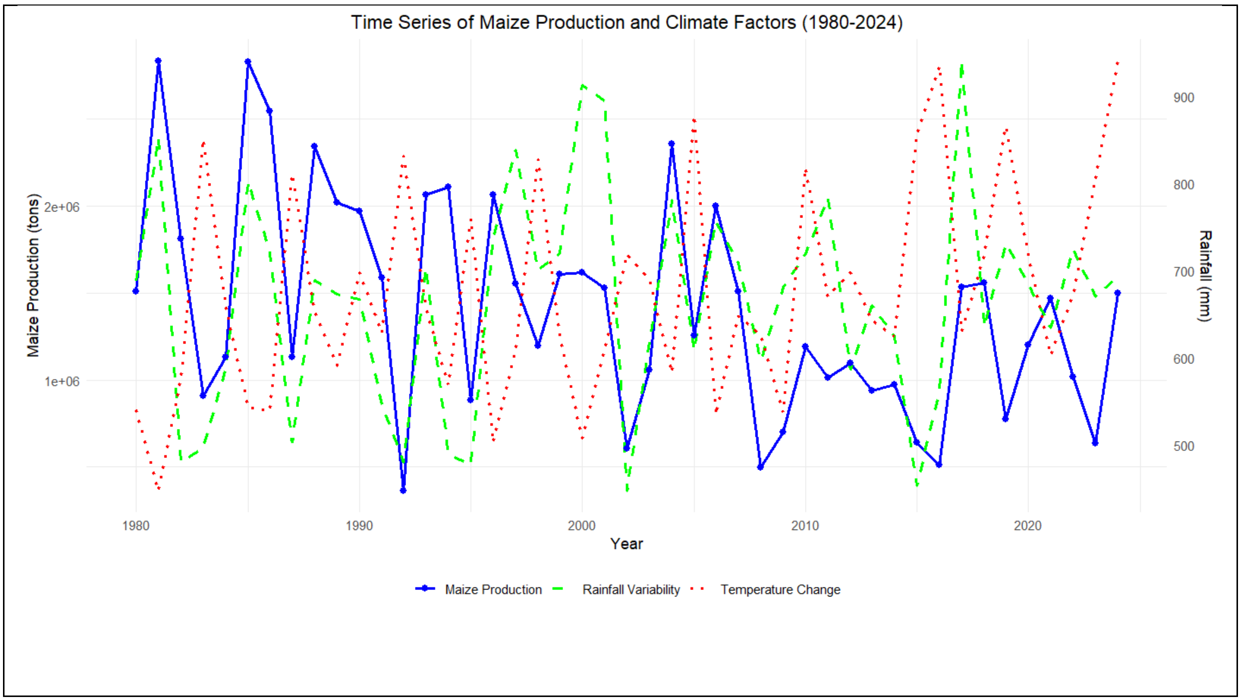


Fig4.4 Maize production, Rainfall variability and Temperature Change Combined Trends

The trend shows that maize production in Zimbabwe closely follows changes in rainfall, with higher maize yields in years with more rainfall and lower yields when rainfall drops. Temperature changes also affect maize, but rainfall has a huge positive impact on production. When rainfall is low, maize production falls sharply, even if temperature changes are not as extreme. This means maize output in Zimbabwe mostly depends on how much rain falls each year, while temperature acts as a control that can make the situation better or worse. The combined plot trend illustration agrees with the findings of Matarira, Unganai and Mukarakate (2019) and Chagutah (2021), who noted that maize yield declines correspond directly with drought years.

Overall, annual trend plots empirically demonstrate the core premise of the Risk and Uncertainty Theory. The unpredictable nature of rainfall shown in your data is exactly the kind of uncertainty that forces farmers to make risk-averse decisions, as described by Moyo et al. (2019).

The observed trends and seasonal fluctuations in both rainfall, temperature changes and maize production may exhibit non-stationarity of the data. Non-stationarity of the data will complicate predictive modelling; hence a stationarity test was done using ADF test in R-Studio.

4.3 Pretest

4.3.1 Stationarity Test

Table 4.2 results of the Augmented Dickey Fuller (ADF) test applied on (Ma)

The Augmented Dickey-Fuller (ADF) test statistics of maize production gave the value of the Dickey-Fuller test statistic as -3.8447, when the lag order was 3 and the p-value was 0.02465 which is less than the level of 0.05. This verifies the alternative hypothesis that the data is stationary.

Table 4.3 outcome. of the Augmented Dickey Fuller (ADF) test applied on (Ra)

The Augmented Dickey-Fuller (ADF) test results to the rainfall variability (RaV) data indicate a Dickey-Fuller statistic of -3.1054, order of lag 3, and p-value of 0.17432199, which is larger than 0.05, the significance level. The outcome verifies the null hypothesis of non-stationarity. The stationarity was attained using differencing.

Table 4.4 The Augmented Dickey-Fuller (ADF) test results for the differenced (diff Ra)

The Augmented Dickey-Fuller (ADF) test repercussions of differenced rainfall variability data (diff Ra) show the Dickey-Fuller test statistic is -5.2057, lag order is 3, and the p-value is 0.01 which is less than 0.05. The implication is the data meet the stationarity criterion.

Table 4.5 results of the Augmented Dickey Fuller (ADF) test applied on (Te)

The Augmented Dickey-Fuller (ADF) test outcome of the temperature time series (Te_ts) is a Dickey-Fuller test statistic of -3.8021, lag order 3, and p-value 0.04293207, which is lower than the 0.05 significance level. This confirms the alternative hypothesis that the data is stationary.

4.3.2 Correlation Test

Fig 4.5 Correlation Matrix

The correlation matrix reveals the relationship between rainfall (Ra), temperature (Te), and maize output (Ma) in Zimbabwe. The positive correlation coefficient value of 0.47 between rainfall and maize output reveals a moderate positive relationship. This shows that moderate to higher levels of rainfall are related to higher maize production. The very high negative relationship of -0.62 between temperature and maize output reveals that higher temperatures reduce maize production. This concurs with empirical studies by Moyo et al. (2022) and Chikwati et al. (2021) that cited rising temperatures and rainfall shortages as the main reasons for low maize yields in Zimbabwe.

4.4 Data Preparation and preprocessing for analytics

4.4.1 Checking for missing values using an R-studio code >(colSums(is.na(DATART))).

Fig 4.6 Results for missing values, using code >(colSums(is.na(DATART)))

Figure 4.6 shows that there were no missing values on the dataset.

4.4.2 Normalisation

Pre-processing was done by normalising the data using the Mini-max normalisation technique in R-studio. The data is normalised between 0 to 1. This will cause all the features to contribute equally to the learning. Normalisation also accelerates convergence during training, reduces exploding and vanishing gradients, and improves the performance of the model overall. Min-max formula used is: $X' = \frac{X - X_{min}}{X_{max} - X_{\min}}$:

Table 4.6 Normalised data from Min-max method of normalization

	Ra	Te	Ma		Ra	Te	Ma
1	0.48769	0.18919	0.46484	24	0.34598	0.49232	0.28197
2	0.82057	0	1	25	0.67623	0.27557	0.80731
3	0.06603	0.26179	0.58526	26	0.32638	0.87281	0.36169
4	0.10192	0.81611	0.22168	27	0.6257	0.17806	0.66175
5	0.28535	0.42872	0.31192	28	0.53	0.407	0.46422
6	0.7189	0.19502	0.99782	29	0.30432	0.36407	0.05426
7	0.55654	0.18707	0.88355	30	0.47324	0.18283	0.1368
8	0.11172	0.73662	0.31112	31	0.55164	0.7504	0.33603
9	0.49095	0.41865	0.80085	32	0.68147	0.45151	0.26242
10	0.45745	0.29147	0.6703	33	0.28128	0.50821	0.29701
11	0.44581	0.50874	0.65128	34	0.43241	0.39534	0.23362
12	0.20149	0.36725	0.49519	35	0.35938	0.35983	0.24805
13	0.06702	0.77954	0	36	0.01332	0.82989	0.11365
14	0.51866	0.4266	0.68829	37	0.23299	0.98728	0.06066
15	0.08568	0.24695	0.70701	38	1	0.37202	0.39275
16	0.06065	0.63222	0.21164	39	0.3811	0.54796	0.48481

17	0.58998	0.11288	0.68924	40	0.57714	0.84738	0.16673
18	0.79381	0.32962	0.48181	41	0.48	0.54743	0.34006
19	0.5132	0.77054	0.33746	42	0.37942	0.31638	0.44822
20	0.55245	0.35877	0.50362	43	0.56297	0.45469	0.26612
21	0.94365	0.12295	0.5089	44	0.45107	0.72761	0.11038
22	0.90715	0.32697	0.47114	45	0.49763	1	0.46049
23	0	0.55008	0.09826				

4.4.3 Feature engineering

The normalised data, as lags and rolling statistics. Feature engineering was employed through the generation of lagged and rolling statistics. Lagged features, such as 2 prior values of a time series, where the model would identify temporal trends and dependencies. Rolling statistics, for instance, rolling means and rolling standard deviations. Rolling statistics helped in smoothing out the volatility of the normalised data, capturing trends and variability over periods.

Fig 4.7 Checking for missing values after feature engineering

R code for removing rows with missing values (NA) from the DATART data frame using na.omit(), after lag and rolling calculations introduced them.

Table 4.7 feature engineering results

Yr	Lag1_Ra	Lag2_Ra	Lag1_Te	Lag2_Te	RollingMean_Ra	RollingSD_Ra	RollingMean_Te	RollingSD_Te
1982	0.82057407	0.487693	0	0.191626	0.4723194	0.6939311	0	0.21849118
1983	0.06602972	0.820574	0.265164	0	0.2744807	0.786883	0.3606707	1
1984	0.10191895	0.06603	0.826624	0.265164	0	0.1843023	0.6073171	0.63268625
1985	0.28535282	0.101919	0.434246	0.826624	0.3348143	0.573976	0.5689024	0.7140901
1986	0.71889802	0.285353	0.197531	0.434246	0.5679584	0.3826658	0.2070122	0.22439896

Table 4.7 above shows the top structure of feature engineering results after the first 2 observations were *NA* due to creation of lag and rolling statistics. The R code in Fig 4.4 omitted the first two rows with non-values. Refer to Appendix 2.

4.5 Model identification

After meeting stationarity test and overseeing missing values, identifying the components of ARIMA (p, d, q) was the next step. To identify the best ARIMA model, Autocorrelation Function and Partial Autocorrelation Function were used.

Fig 4.8 ACF and PACF plots for model identification.

The PACF plot of the differenced maize yield has a spike at lag 0. This is a sign of Moving Average p=0 that the series have been differenced, The PACF plot, there are no spikes. This is a sign of the autoregressive (AR) component, q=0. The behaviour demonstrated that, differencing d=1 stabilized the series sufficiently. This resulted in Arima (0_1_0) being the suggested model.

4.6 PARAMETER ESTIMATION

Table 4.8 Parameter Estimation

"Best ARIMA Model from auto. Arima:"
Regression with ARIMA (0,1,0) errors
Coefficients: train_rainfall train temperature
0.1529 -0.4753
Standard Error 0.0953 0.1261
sigma^2 = 0.05831: log likelihood = 1.1
AIC=3.8 AICc=4.66 BIC=8.2

The parameter estimation provides a meaningful from the observation's components of the best

ARIMA (0,1,0) model. In this model AR (0) and MA (0), indicates no autoregressive or moving average components are necessary. The differencing component (d=1) is pivotal, as it helped in achieving stationarity. (Refer to appendix 1)

4.7 Model diagnostics

4.7.1 Normality test

Table 4.9 Shapiro wilk test

This image cannot currently be displayed.

Table 4.9 shows the result from The Shapiro-Wilk test of normality for residuals, which are differences between observed and fitted values in the model. The result shows a p-value of 0.1666868 that is much greater than the typical cutoff of 0.05. The result verifies the normality assumption of residuals.

4.7.2 Homoscedasticity test

Table 4.10 Breusch Pagan test

This image cannot currently be displayed.

As the p-value (0.2047737) is greater than 0.05, we can state that there is no indicative proof of heteroscedasticity in model's residuals. This means that the residuals are of fixed variance, which is a welcome aspect of regression models.

4.7.3 Independence test

Fig 4.9 ACF and PACF

The graphs provided, Figure 4.6 ACF and PACF, plot the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of residuals. It is seen that ACF and PACF correlation bars fall inside the blue dotted lines of confidence interval and beyond lag 1. It shows no autocorrelations, and hence residuals are independent.

4.8 Forecasting ARIMA testing set (0_1_0)

Table 4.11 predicted values

Year	Actual	ARIMA
2016	0.47471	0.11606
2017	0.48587	0.52941
2018	0.1671	0.22325
2019	0.27125	0.18741
2020	0.3408	0.30331
2021	0.4492	0.42658
2022	0.2667	0.39591
2023	0.11062	0.20451
2024	0.4615	0.08278

The expected maize yield from the test set,2016 to 2024 shows an improvement in 2017 due to favourable rainfall and temperature. Fluctuations later depict the negative impact of rising temperatures and erratic rainfall patterns, leading to uncertainty and reduced maize production forecast by 2024.

Fig 4.10 forecasted plot for ARIMA (0_1_0)

Fig 4.10 plots actual and forecast maize production between 2016 and 2024. The ARIMA model does trend but struggles with precision especially during peak years, as in the year 2017, production surpassed expectations. The model uses the overall trends but was unable to pick up the trends during times of high variability and fluctuations.

4.9 Long Short-Term Memory Model Building

4.9.1 Hyperparameter tuning.

After feature engineering, Hyperparameter tuning optimized the parameters of LSTM model to achieve maximum performance on prediction. Some major hyperparameters such as number of hidden neurons, learning rate, and batch size impacted how well the model learns from historical data and generalizes future observations.

TABLE 4.12 different hyperparameter settings and RMSE

Size	Maxit(epoxy)	Learn_rate	RMSE
10	100	0.01	0.07517378
10	50	0.01	0.25179
25	50	0.01	0.28584
50	50	0.01	0.31981
10	100	0.01	0.26014
25	100	0.01	0.25909
50	100	0.01	0.3027
10	50	0.05	0.3146
25	50	0.05	0.34404
50	50	0.05	0.34968

10	100	0.05	0.26522
25	100	0.05	0.33573
50	100	0.05	0.41391

Table 4.12 gives an analysis of different hyperparameter tuning in achieving optimal performance with the best Root Mean Squared Error. The best model configuration identified for the LSTM model is highlighted in green. 10 hidden neurons show a simple architecture which balances complexity and the model's capability to acquire from the data without overfitting. Set at 0.01 learning rate is moderately high, allowing the model to make substantial updates to weights during training. This facilitates faster convergence. 100 epochs were conducted in the training process, providing sufficient iterations for the model to learn from the training data. This duration allows the model to refine its weights and improve prediction accuracy effectively. The lowest RMSE of **0.07517378**, indicating this configuration is the most effective for minimizing prediction error.

Figure 4.11 Lstm Model Structure

Figure 4.11 displays the printed summary of an LSTM (Long Short-term Memory) model object in Specifically from the RSNNS package. The model is of the "Elman" network type, i.e., a kind of recurrent neural network. The summary shows that the given model has 6 input features and 1 output, was trained for up to 100 iterations, uses specific initialization and learning functions with their associated parameters (0.01 learning rate), The architecture consists of 10 hidden units, as indicated by size 10. It also lists a number of internal components and parameters of the model, such as the number of inputs, outputs, initialization and learning functions, and error calculation settings. The output provides a technical explanation of the structure and setup of the neural network.

4.9.2 Forecasting LSTM model.

Table 4.13 Predicted Vs Actual: Test set(20%)

Year	Actual	Predicted
2016	0.47471	0.56574
2017	0.48587	0.51923
2018	0.1671	0.18968
2019	0.27125	0.29864
2020	0.3408	0.36314
2021	0.4492	0.38863
2022	0.2667	0.24864
2023	0.11062	0.18937
2024	0.4615	0.28973

Table 4.13 Lstm plot between predicted and actual values was attained after smoothing. The process enhanced visualisation to facilitate easier comparison between the actual and predicted values. Interpretation of the plot was also improved.

Fig 4.12 LSTM plot for predicted vs actual values from the Test set(20%)

Figure 4.12 shows the actual and predicted maize production values as derived from the LSTM model, from test set (2016 to 2024). 2016 to 2018, an increase of actual and predicted values of production is observed, which shows a period of growth. 2019, shows a dip in both lines. This shows the LSTM predictions very closely resemble the actual values.

4.10 Model Evaluation and Comparison

4.10.1 Performance Matrices

Table 4.14 performance matrices from ARIM and LSTM

METRICS	ARIMA (0_1_0)	LSTM
MAPE	44.45307 %	20.54143%
MAE	0.1172049	0.05842778
MSE	0.027365	0.005651098
RMSE	0.1654237	0.07517378
R-squared	0.184202	0.6793316

A combination of the matrices in table 4.14 provide a comprehensive evaluation of model performances, considering both the magnitude and direction of errors (Wilson, 2022). The performance metrics for the ARIMA and LSTM models highlight a difference in their predictive accuracy regarding maize production. The ARIMA model shows a high MAPE of 44.45307%, indicating poor forecast accuracy, while the LSTM model achieves a much lower MAPE of 20.54143%, suggesting it performs better. Additionally, the MAE and RMSE values for LSTM (0.0623 and 0.07517378, respectively) are lower than those for ARIMA (0.1172049 and 0.1654237), confirming that LSTM predictions are much closer to actual values. The R-squared value of 0.6793316 for the LSTM model indicates that approximately 67.9% of the variance in maize production is explained by the model, proving strong explanatory power. In contrast, ARIMA's performance metrics suggest it failed to capture the trends in maize production making LSTM the preferred model for accurate forecasting.

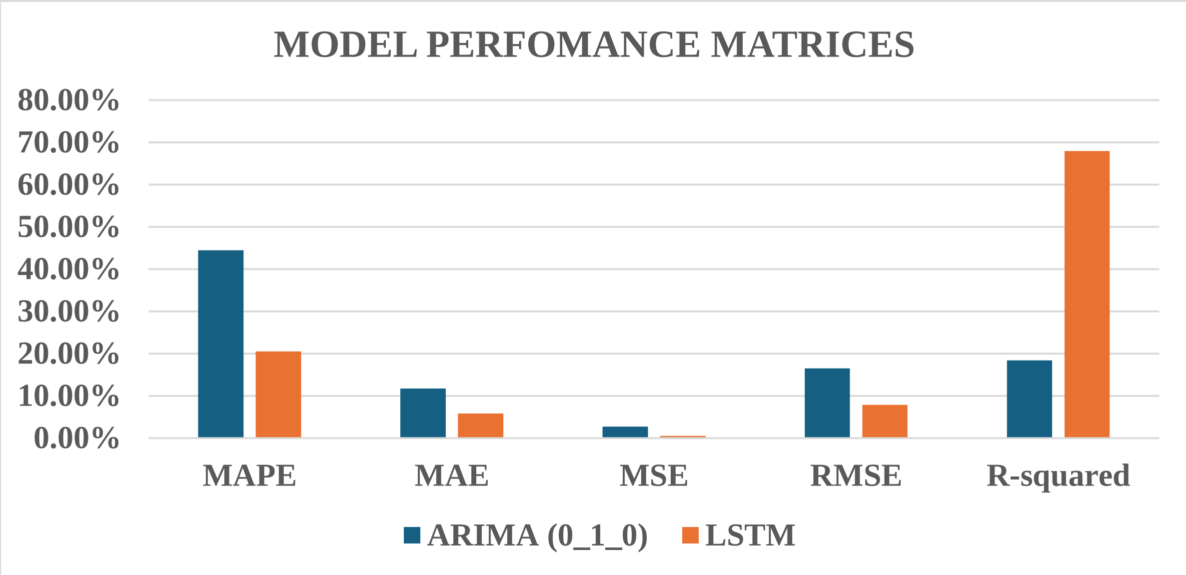


Fig 4.13 Comparison Bar Graph for performance.

Fig 4.13 presents the performance metrics of ARIMA and LSTM models to predict maize production and signifies that LSTM possesses larger performance metrics compared to ARIMA on all the metrics, i.e., MAPE, MAE, MSE, RMSE, and R-squared. LSTM possesses smaller error rates and a better fit for data, and therefore it is the optimum model to be utilized for accurate forecasting.

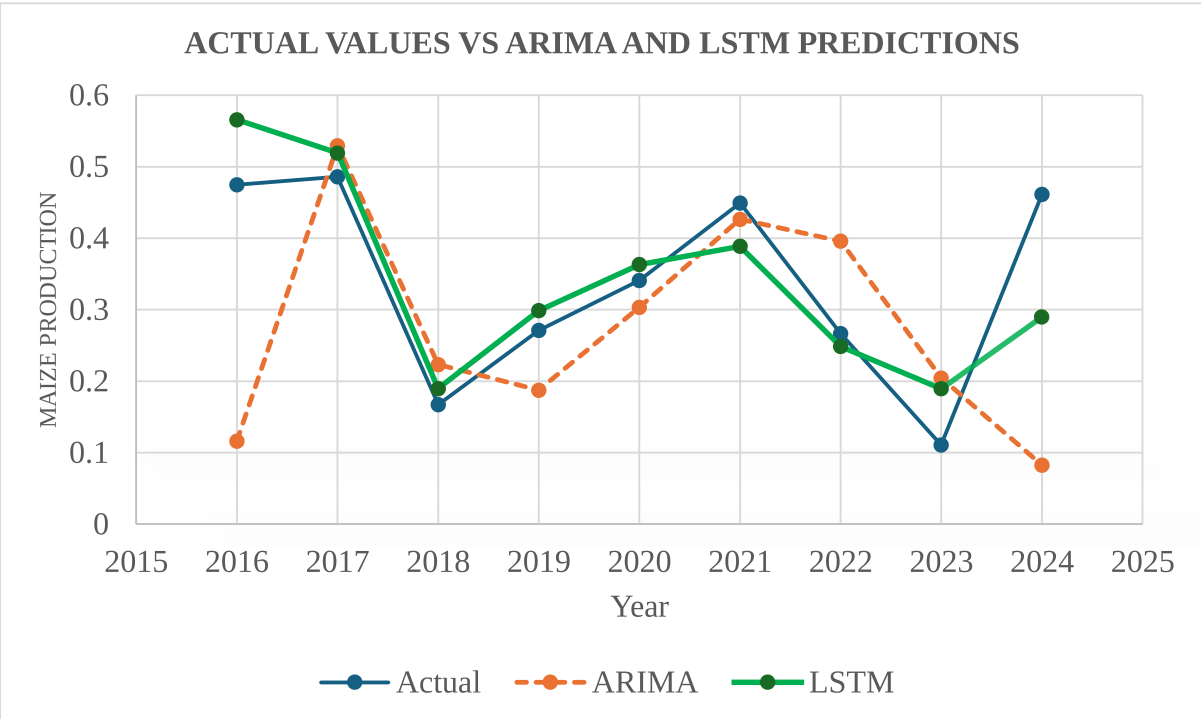


Figure 4.14 Line plot: ARIMA and LSTM prediction performance.

Figure 4.14 illustrates a comparative model of maize production forecasts based on ARIMA and ANN models with the actual observed values between the years 2016 to 2024, (Hyndman and Athanasopoulos, 2018). The ARIMA model, a conventional time series approach, indicates and alignment in short term intervals and high deviation on long term intervals from real maize production values. This is indicative of its inability to follow fluctuations in times of high variability. On the contrary, the LSTM model, which is capable of identifying complex patterns and long-distance dependencies, compares very much better with the real data (Mahaluça. F et all). The study brings Schultz's Theory of Agricultural Transformation into existence by demonstrating that employing superior technology, such as Artificial Neural networks, LSTM models, can actually enhance farming. Instead of relying with older methods like ARIMA, the research used a smarter tool that gives more accurate results. This aids agricultural team and decision makers to plan better and use resources wisely, as suggested by Schultz. By recommending that stakeholders use advanced tools like artificial neural networks, your study shows a clear way to modernize agriculture and achieve real progress.

4.11 Maize Production Outlook For 2025-2029

Maize was predicted using the best Predictive Time series model, the LSTM model. This was done based on the predictive accuracy of this model using the test set, which is 20% of data. The LSTM model performed better than the ARIMA model.

Maize production in (00,000,000) metric tonnes

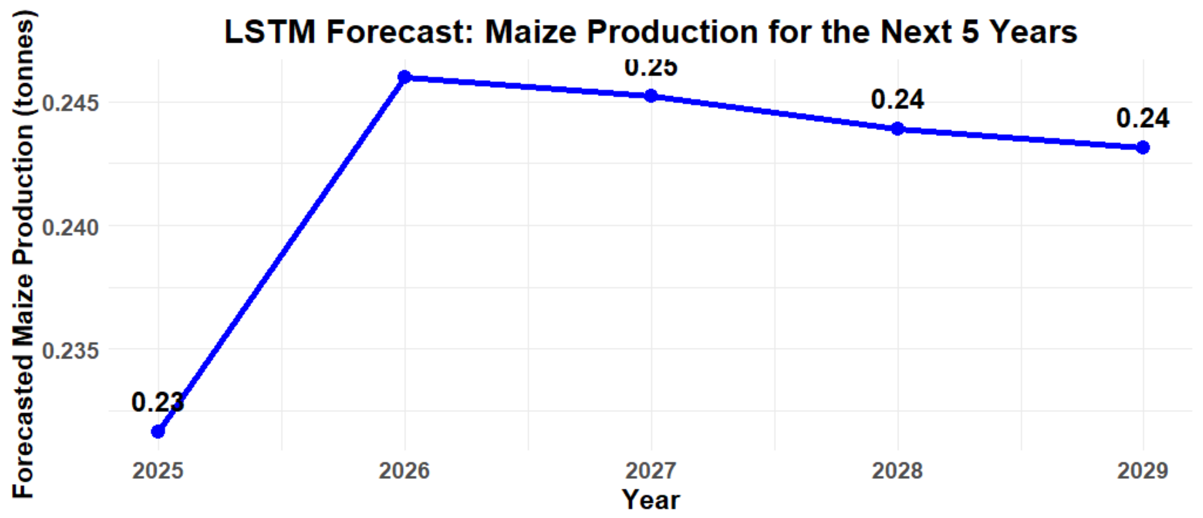


Fig 4.15 Maize Production Forecast For 2025-2029

The LSTM forecast for maize production in Zimbabwe from 2025 to 2029 indicates a slight increase in production, with values projected to stabilize around 2.4 million tonnes. The peak in 2026 may correspond to favourable climatic conditions, while the gradual decline in later years could indicate drought stress or temperature shifts. The forecast also suggests a modest recovery or maintenance of maize yields in the coming years. The forecast results reflect resilience in production despite potential climatic challenges. The inclusion of rainfall variability and temperature change as influencing factors shows the importance of these variables in determining maize output.

4.12 Discussion of Findings

The findings of this study show the relationship of rainfall variability and temperature changes on maize production in Zimbabwe. The time series analysis illustrated fluctuations in maize yields, with rainfall serving as a primary driver and temperature acting as a moderating factor. The descriptive statistics confirmed that while maize production averages around 1.41 million tons, the variability in yields is substantial and displays the need for employing advanced predictive models. The correlation analysis verifies a moderate positive relationship between rainfall and maize production, while temperature exhibits a strong negative correlation, indicating its adverse effects on crop yields. The test for ADF indicated that the data generated from maize production were stationary, which is relevant for an accurate forecast.

Comparative analysis of the ARIMA and LSTM model reveals that LSTM has captured all types of complex patterns long dependencies in maize yield forecasting. The 2.3million metric tons of maize production resulted assuming the forecast plot of maize produced for Zimbabwe in 2025 (retrieved from newZWire website 29/05/ 2025). This is the effectiveness of Long Short-Term Memory. The base ARIMA model, being restrictive due to linearity, fails to explain non-linear interactions between climatic variables and agricultural output. In contrast, the LSTM model showcases its strength in terms of predictive-accuracy, which is seen through lower MAPE, MAE, RMSE, and R-squared values.

The projected slight growth of maize production from 2025 to 2026 and then a slow to stable decline between the years 2026 and 2029 is worrisome for food security and agricultural sustainability. FAO (2022) mentioned that maize production will be lower in subsequent years because of climate variability. This trend requires initiative-taking interventions, in the form of adopting climate-resilient farming practices, irrigation systems, and integration of advanced predictive models for yield forecasts. The study highlights this need urgency through policy advocacy, financial planning, and education measures to arrange and prepare tools at the farmer's end about risk mitigation related to climate change.

4.13 Summary

Hypothesis tests were conducted for the models to be fitted, and they were well presented in this chapter. The ARIMA model selection process identifies ARIMA (0,1,0) as the best-fitting model based on AIC and BIC values. However, performance metrics reveal its limitations in capturing complex climatic interactions. Artificial Neural Network, The LSTM model, leveraging deep learning techniques, outperforms ARIMA, demonstrating superior accuracy in maize production forecasting.

The results of the paper indicate that rainfall is in explaining maize yields and had been negatively affected by temperature. The trend of production of maize and the prospects of shortfalls calls for interventions geared toward food security. This chapter becomes an entry point into the conclusion and recommendation phase.

CHAPTER 5: FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

5.0 Introduction

This chapter brings together the research findings, drawing conclusions from data analysis and providing practical recommendations to agricultural stakeholders in Zimbabwe. What a changing rainfall variability and temperature influence maize production and how well-advanced predictive models like LSTM perform, is explored.

5.1 Summary of Findings

The study confirms that rainfall variability has had a serious impact on maize production in Zimbabwe, with temperature as a moderating factor. The correlation analysis first verifies a moderate positive relationship (0.47) between rainfall and maize yields but temperature change as a control variable exhibits a very strong negative correlation (-0.62). These correlations are stating that during the period of high to moderate rainfall production of maize should be increased and during the period of high temperature is causing a decline in maize production.

Comparative analysis between ARIMA and ANN models illustrates that ANN performs better amongst these models as it captures more quantity of non-linear interactions. An Artificial Neural Network, Long Short-Term Memory model was superior to the ARIMA model in yielding lesser MAPE (20.54%), MAE(0.058), RMSE(0.075) and maximum R-squared (67.9%) depicting it as effective in forecasting agricultural outputs under variable climatic conditions. The above aligns with literature that insisted that ANNs are superior to managing non-linear relations and intricate interactions (Breiman, 2001; Mugabe et al., 2021). LSTM model's superiority was confirmed by a lower MAPE (20.54%) and a strong R-squared (0.679) being supported by Zimbabwean literature. This finding directly addresses the research gap you identified, offering a clear answer as to which model is more effective for this type of agricultural forecasting. The research verifies the recommendations from a study by Mapuwei, et al. (2022), who used ARIMA for tobacco, and recommended that other academics should explore ANN models for comparison.

The last objective was attained through the forecasted decline in maize production from 2025 to 2029 using the best-chosen ANN. The decreasing trend in maize production indicates the need for urgency of implementing adaptive agricultural strategies. The Zimbabwean government should continue assisting farmers via Command agriculture called Pfumvunza. Enhanced irrigation infrastructure also plays an important role in stabilizing maize yields. The study also research explains the necessity of incorporating climate information into farming planning, and it suggests policy-led solutions to address climate risks. The LSTM forecast plot projects the production of maize in Zimbabwe between the years 2025 and 2029. It indicates an increase from 2.3million tonnes in 2025 to a high of approximately 2.5million tonnes in 2026, then a steady and slow decrease, settling at 0.24 tonnes between 2027 and 2029. This indicates moderate growth and then consistent yields of maize for the next five years, provided that recent tendencies in rainfall, temperature, and other factors modelled remain constant.

5.2 Conclusion

The study concludes that climate variability, i.e., rainfall pattern and temperature variations, is a serious threat to maize production in Zimbabwe. The study indicates that an advanced prediction methodology needs to be adopted for precipitation of high accuracy in forecasting. The study highlights that Zimbabwe needs to lead the way to enhance agricultural systems in order to increase the production of maize. Additionally, the study demands the significance of incorporating climate information into farm planning for ensuring sustainable food production.

5.3 Recommendations

a. Use of ANN (Advanced Predictive Models) in Forecasting

Stakeholders are encouraged to include advanced predictive models, Artificial Neural Networks (ANN), in the forecasting. This is because of their super-power in the forecasting of agricultural yield under climatic conditions. (Zhang, (2022).)

These predictive models will better understand the intricacies of agricultural data impacted by climatic factors and help in better decision-making and planning.

b. **Maize Production Projection Based on Rainfall Patterns and Change in Temperature**.

According to the LSTM projection of a moderate growth in Zimbabwe's maize production until 2026 and then leveling off, therefore, strategic advice is important. For maximizing the initial growth, short-term interventions should be aimed at securing farmers' access to improved inputs, extension services for optimal practices, and irrigation infrastructure. To address the plateau anticipated from 2027, it is important to invest more in research and development for new maize varieties, wider implementation of climate-smart agriculture, and improvement of soil health. Enhancing climate resilience through crop diversification, drought preparedness plans, and early warning systems from is vital to mitigate the impact of rainfall variability and temperature changes. Reducing post-harvest losses by training farmers on good grain handling and better storage facilities and enhanced market linkages will ensure availability of food and reasonable prices for farmers. Lastly, a strong data collection framework to monitor yields, climate factors, and farming practices, combined with periodic updating of the LSTM model, will facilitate data-based decision-making and ongoing improvement. It will be designed to attain a resilient and sustainable maize production system, ensuring food security, improved livelihoods for agricultural communities, and more resilience along the value chain for Zimbabwe. Implementing these strategies needs to be done in May 2025, readying the plant for the upcoming planting seasons.

c. Education and Training

In modern agricultural development, a focus on education and training programs is critical for attaining sustainable growth and improvement in farming communities' resilience. Targeted interventions, such as strengthening agricultural extension services through incorporating Agritex officer known as Madhumeni, promoting farmer field schools such as Blackfordby College of Agriculture, and Harare Technical and Agriculture. Moreso, enhancing vocational training, are crucial for disseminating knowledge, empowering farmers, and creating a skilled workforce (Anderson & Feder, 2007). Including climate change education and encouraging digital literacy are also critical to empowering farmers with the knowledge necessary to adapt to new environmental issues and take advantage of technological progress (FAO, 2016). Proper investments in training and education are crucial to maximizing the agricultural sector's potential and protecting food security under a changing world.

d. Policy Advocacy

Policy advocacy initiatives must focus on advancing climate-smart agriculture and sustainable water resource management. This also includes policy advocacy to encourage the adoption of drought-resistant varieties, appropriate irrigation techniques, and investments in water harvesting infrastructure. This response will assist in directing the National Agriculture Policy Framework (NAPF) during the realignment 2019-2030 for the overall purpose of providing policy direction and guidance on how to enhance and facilitate the sustainable entry of domestic and foreign investment and resources to change the trajectory of the agricultural sector through enhanced and sustained agricultural growth, production, productivity, and competitiveness.

e. Financial Planning

Command agriculture, a Zimbabwean government programme which provides loans to farmers for inputs like seeds, fertilisers fuel and other must be implement well. This will help to increase agricultural production, improve food security, and reduce imports cost. Moreso, more financial institutions in Zimbabwe like AgriBank should be developed tailored financial products that support farmers in investing in adaptive technologies and improved practices. Increase access to credit and insurance schemes can provide a safety net against the uncertainties of climate impacts. Financial institutions should develop loan products tailored for smallholders and facilitate mobile banking services to enhance access to services in rural areas

5.4 Areas for Further Research

This study creates a way for other future studies. Besides focusing on maize production based on climatic factors, further research should focus on other factors that include socio-economic factors and political factors on other crops like bananas in Zimbabwe. Moreso employing other advanced predicted models that include ensembled machines like, Random Forest Gradient boosting, other artificial neural networks like Feedforward Neural Networks models and other machine learning predictive tool. The research also paves a way for further studies to focus on large dataset that enables the models to read trends. Studies examining the effectiveness of specific adaptive strategies in various regions of Zimbabwe will also be valuable.

5.5 Summary of Chapter

In summary, this chapter highlights the findings regarding the effects of rainfall variability and temperature changes serving as a control variable on maize production in Zimbabwe. Research objectives and questions were satisfactorily answered in the chapter. Suggested recommendations aim to guide stakeholders on how to respond to climate variability issues, so as to ensure sustainable agriculture and preserve food security in the future.

REFERENCES

1. Abaye, T. et al. (2019) 'Rainfall variability and maize production in Kenya', Journal of Agricultural Science, 17(2), pp. 123-135. Available at: https://maxapress.com/article/doi/10.48130/tia-0024-0030
2. Anderson, J. R., & Feder, G. (2007). Agricultural extension. In R. Evenson & P. Pingali Eds.), *Handbook of Agricultural Economics* (Vol. 3, pp. 2343-2378). Elsevier. Available at: https://www.sciencedirect.com/science/article/pii/S1574007206030441
3. Awe, O., Okeyinka, A., & Fatokun, J. O. (2020). An alternative algorithm for ARIMA model selection. 2020 international conference in mathematics, computer engineering and computer science (ICMCECS), (pp. 1– Available at: https://ieeexplore.ieee.org/abstract/document/9077654/
4. Bhimavarapu, U., Battineni, G. and Chintalapudi, N., 2023. Improved optimization algorithm in LSTM to predict crop yield. *Computers*, *12*(1), p.10. Available at: https://www.mdpi.com/2073-431X/12/1/10
5. Box, G. E. P. and Jenkins, G. M. (1976) Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day. https://cir.nii.ac.jp/crid/1130000797820830336
6. Breiman, L. (2001) 'Random Forests', Machine Learning, 45(1), pp. 5–32. doi: 10.1023/A:1010933404324. Available at:
7. https://link.springer.com/article/10.1023/a:1010933404324
8. Gwenzi, J., Mafongoya, P.L. and Mashonjowa, E., 2022. Effects of climate change on food production in semi-arid areas: a case study of Uzumba Maramba Pfungwe district, Zimbabwe. In *Handbook of Climate Change Across the Food Supply Chain* (pp. 451-465). Cham: Springer International Publishing. Available at: https://link.springer.com/chapter/10.1007/978-3-030-87934-1_26
9. Chen, C.-H.;Lai, J.-P.; Chang, Y.-M.; Lai, C.-J.; Pai, P.-F. A study of optimisation in deep neural networks for regression. Electronics 2023, 12, 307 Available at: https://dspace.unila.edu.br/items/cf231d6b-8ca7-47e1-bbee-413136c9358f
10. De Myttenaere, A., Golden, B., Le Grand, B. and Rossi, F., 2016. Mean absolute percentage error for regression models. *Neurocomputing*, *192*, pp.38-48. Available at: https://www.sciencedirect.com/science/article/pii/S0925231216003325
11. ENAYABA, O., AJIBOYE, A., EGWU, L., AGBAMS, P. and DAMOYE, T., 2024. "TIME SERIES, MACHINE LEARNING AND ENSEMBLE MODELS TO FORECAST RICE CROP YIELD IN NIGERIA UNDER CLIMATIC VARIABILITY."

("View of TIME SERIES, MACHINE LEARNING AND ENSEMBLE MODELS TO FORECAST ...") *International Journal of Agricultural Research and Biotechnology*. Available at: https://taapublications.com/tijarbt/article/view/269

12. FAO (Food and Agriculture Organization) (2020) The Impact of Climate Change on Agriculture. Rome: FAO. Available at: https://www.taylorfrancis.com/chapters/edit/10.4324/9781003179900-41/food-agriculture-organization-united-nations%E2%80%94fao-helen-canton
13. FAO. (2016). *Climate-smart agriculture sourcebook*. Food and Agriculture Organization of the United Nations. Available at: https://www.cabidigitallibrary.org/doi/full/10.5555/20163090494
14. Gao, J., 2024. R-Squared (R2)–How much variation is explained?. *Research Methods in Medicine & Health Sciences*, *5*(4), pp.104-109. Available at: https://journals.sagepub.com/doi/abs/10.1177/26320843231186398
15. Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). "LSTM: A Search Space Odyssey." IEEE Transactions on Neural Networks and Learn ing Systems, 28(10), 2222-2232. Available at: https://doi.org/10.1109/TNNLS.2016.2582924
16. Gwimbi, P., 2009. Cotton farmers' vulnerability to climate change in Gokwe District (Zimbabwe): impact and influencing factors. *JAMBA: Journal of Disaster risk studies*, *2*(2), pp.81-92. Available at: https://journals.co.za/doi/abs/10.10520/EJC51166
17. Mupangwa, W., Chipindu, L., Nyagumbo, I., Mkuhlani, S. and Sisito, G., 2020. ("An Approach for Potato Yield Prediction Using Machine Learning ...") Evaluating machine learning algorithms for predicting maize yield under conservation agriculture in Eastern and Southern Africa. *SN applied sciences*, *2*(5), p.952.Heale, R. and Twycross, A., 2018. What is a case study?. *Evidence-based nursing*, *21*(1), pp.7-8. Ali, Z. and Bhaskar, S.B., 2016. Basic statistical tools in research and data analysis. *Indian journal of anaesthesia*, *60*(9), pp.662-669. Available at: https://link.springer.com/chapter/10.1007/978-981-19-1412-6_27
18. Hochreiter, S.and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780. Available at: https://ieeexplore.ieee.org/abstract/document/6795963/

19. Hodson, T.O., 2022. Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. Geoscientific Model Development Discussions, 2022, pp.1-10. Available at: https://gmd.copernicus.org/preprints/gmd-2022-64/

20. Hughes, M., 2015. Analysis of notation data: reliability. In *Essentials of Performance Analysis in Sport* (pp. 169-179). Routledge. Available at: https://www.taylorfrancis.com/chapters/edit/10.4324/9781315776743-11/analysis-notation-data-mike-hughes

21. Hyndman, R. J. and Athanasopoulos, G. (2018) Forecasting: Principles and Practice. 2nd edn. Melbourne, Australia: OTexts. Available at: https://otexts.com/fpp2/ (Accessed: 12 October 2024).

22. Intergovernmental Panel on Climate Change (IPCC) (2021) Climate Change 2021: The Physical Science Basis. Available at: https://www.ipcc.ch/report/ar6/wg1/ (Accessed: 12 October 2024).

23. IPCC (Intergovernmental Panel on Climate Change) (2013) Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Available at: https://cir.nii.ac.jp/crid/1371413280484207233

24. Kamilaris, A., & Prenafeta-Boldú, F. X. (2022). Deep Learning in Agriculture: A Review. *Sensors*, 22(1), 0101. Available at: https://www.researchgate.net/profile/Milon-Biswas-2/publication/359243072_Deep_Learning_in_Agriculture_A_Review/links/6230cc1c4ba65b248132df40/Deep-Learning-in-Agriculture-A-Review.pdf?_sg%5B0%5D=started_experiment_milestone&origin=journalDetail&_rtd=e30%3D

25. Kumar, V., & Garg, M. L. (2018). Predictive analytics: a review of trends and techniques. *International Journal of Computer Applications, 182*, 31–37. Available at: https://www.researchgate.net/profile/Vaibhav-Kumar-16/publication/326435728_Predictive_Analytics_A_Review_of_Trends_and_Techniques/links/5c484f6692851c22a38a6027/Predictive-Analytics-A-Review-of-Trends-and-Techniques.pdf

26. .Mackey, A., & M, S. (2023). Research Methods in Second Language Acquisition. *The Morden Language Journal.* Available at: *https://books.google.com/books?hl=en&lr=&id=h8LTZ4eyF7QC&oi=fnd&pg=PR7&dq=26.%09.Mackey,+A.,+%26+M,+S.+(2023).+Research+Methods+in+Second+*

Language+Acquisition.+The+Morden+Language+Journal.&ots=i8W3GYt3EY&sig=tP6EqclYVxvKsyRvuN75O2HHRIM

27. Mahaluça, F., Carsane, F. and Vilanculos, A., 2024. "Forecasting Maize Production in Mozambique: A Comparative Analysis of Arima and Lstm Models." ("ISSN: 2771-7763 Journal of Economic Research & Reviews") *J Eco Res & Rev*, *4*(3), pp.01-15. Available at: https://www.opastpublishers.com/citation/forecasting-maize-production-in-mozambique-a-comparative-analysis-of-arima-and-lstm-models-8288.html
28. Mapuwei, T. W., Ndava, J., Kachaka, M., & Kusotera, B. (2022). An Application of Time Series ARIMA Forecasting Model for Predicting Tobacco Production in Zimbabwe. *American Journal of Modeling and Optimization, 9*, 15–22. Available at: https://www.researchgate.net/profile/Tw-Mapuwei/publication/366465209_An_Application_of_Time_Series_ARIMA_Forecasting_Model_for_Predicting_Tobacco_Production_in_Zimbabwe/links/63a2d3899835ef259039770d/An-Application-of-Time-Series-ARIMA-Forecasting-Model-for-Predicting-Tobacco-Production-in-Zimbabwe.pdf
29. Matarira, C. H., Unganai, L. and Mukarakate, M. (2019) 'The impact of climate change on rainfall patterns and maize production in Zimbabwe', Zimbabwe Journal of Agricultural Research, 12(4), pp. 256–277. Available at: https://www.researchgate.net/publication/387402662_Impact_of_Rainfall_and_Temperature_as_Aspects_of_Climate_Change_on_Maize_Production_in_Zimbabwe
30. Moyo, P., Nyanga, P. H., & Mhlanga, N. (2016). Time series analysis of rainfall and maize yields in Zimbabwe: Evidence from Midlands Province. International Journal of Agricultural Statistics, 14(2), pp. 23-31. Available at: https://edepot.wur.nl/495461
31. Mugabe, F. T., Matarira, C. H. and Unganai, L. (2021) 'Climate change and agricultural yields in Zimbabwe: The role of rainfall patterns in maize production', Zimbabwe Agricultural Journal, 45(2), pp. 78–95. Available at: https://link.springer.com/chapter/10.1007/978-981-97-2155-9_3
32. Mugabe, T., Matarira, C., & Unganai, L. (2021). Predicting maize yield in Zimbabwe using artificial neural networks with climatic variables. Journal of Agricultural Science and Technology, 23(3), pp. 34-45. Available at: https://ieomsociety.org/proceedings/2023vietnam/46.pdf
33. Mupangwa, W. et al. (2017) 'Rainfall variability and maize production in Zimbabwe', Journal of Agricultural Science, 15(1), pp. 1-12. Available at: https://www.sciencedirect.com/science/article/pii/S0167880920300839

34. Mupangwa, W., Chaguma, W., & Mushore, T. (2017). Time series analysis of rainfall-maize yield relationships in Zimbabwe. Journal of Agricultural Science, 14(2), 1-10. Available at: https://www.aimspress.com/article/10.3934/agrfood.2016.1.85
35. Mutasa, F. and Nyambara, P. (2020) 'Rainfall variability and agricultural performance in Zimbabwe: A time series analysis', Zimbabwe Journal of Economics and Agricultural Development, 13(1), pp. 145–162.
36. Mutasa, M., & Nyambara, T. (2020). Rainfall Variability and Maize Production: A Case of Smallholder Farmers in Zimbabwe. Journal of Climate and Agricultural Research, 35(4), pp. 78-89.
37. Namin, S.S. and Namin, A.S., 2018. Forecasting economic and financial time series: Arima vs. LSTM. *arXiv preprint arXiv:1803.06386* *https://arxiv.org/abs/1803.06386*
38. Patterson J., 2017. Deep Learning: A Practitioner's Approach, O'Reilly Media. https://books.google.com/books?hl=en&lr=&id=qrcuDwAAQBAJ&oi=fnd&pg=PR2&dq=38.%09Patterson+J.,+2017.+Deep+Learning:+A+Practitioner%E2%80%99s+Approach,+O%E2%80%99Reilly+Media.&ots=6onEsBgpPS&sig=OCOAa00KPm3pf82v7pp08grYvTY
39. QuantInsti (n.d.) Understanding stationarity in time series data. Available at: *https://blog.quantinsti.com/stationarity* (Accessed: 14 March 2025).
40. Rathod, S., & Mishra, G. C. (2018). Statistical models for forecasting mango and banana yield in Karnataka, India. ("Statistical Models for Forecasting Mango and Banana Yield of ... - SID") Journal of Agricultural Science and Technology, 20(4), 803–816. Retrieved from *http://www.jast.modares.ac.ir*
41. Riya, T. (2023). Manuscriot and Grants, Reporting Research:enago academy. Trget population and sample definitions. Available at: https://www.google.com/search?sca_esv=c5a1f40b3c8677e8&rlz=1C1CHBF_enZW1141ZW1141&sxsrf=AE3TifMpkP7D5SIGEiHhLJjj8zUnDJxBKw:1751413506759&q=41.+Riya,+T.+(2023).+Manuscript+and+Grants,+Reporting+Research:enago+academy.+Target+population+and+sample+definitions.&spell=1&sa=X&ved=2ahUKEwiFlbPf65yOAxUlR_EDHY_pO6EQkeECKAB6BAgMEAE
42. Rurinda, J., Mapfumo, P., Van Wijk, M.T., Mtambanengwe, F., Rufino, M.C., Chikowo, R. and Giller, K.E., 2014. "Comparative assessment of maize, finger millet and sorghum for household food security in the face of increasing climatic risk." ("Comparative assessment of maize, finger millet and sorghum for ...") *European Journal of Agronomy*, *55*, pp.29-41. Available at:

https://www.sciencedirect.com/science/article/pii/S1161030114000021

43. Taherdoost, H., 2021. "Data collection methods and tools for research; a step-by-step guide to choose data collection technique for academic and business research projects." (“(PDF) Data Collection Methods and Tools for Research; A Step-by-Step ...”) *International Journal of Academic Research in Management (IJARM)*, *10*(1), pp.10-38. Available at: https://hal.science/Hal-03741847/
44. Wilson, D. (2022). Comprehensive Model Evaluation. Applied Statistics, 58(4), 567-589. https://pubs.acs.org/doi/abs/10.1021/acsami.2c15095
45. Wyk.(2012).Retrievedfrom: Available at: https://s3.amazonaws.com/academia.edu.documents/4322604 1/Research_and_Design-I1.pdP?response-content-disposition -inline%63 B% 20fiename% 3DResearch design _ and_ methods_ Part_I pdf & X-Amz Algorithm=AWS4-HMAC-SHA256&X-Amz CredentialAKIAIWOWYYGZ2Y53UL3A%2+2019081
46. Zhang, Y., & Chen, X. (2022). Application of LSTM Neural Networks in Time Series Forecasting of Agricultural Yield. *International Journal of Agricultural Research*, 17(2), 101-110. Available at: https://www.mdpi.com/2071-1050/14/17/10483
47. Zimbabwe Meteorological Services Department (2023) Annual Climate Report: Rainfall and Climate Trends in Zimbabwe, 1980-2023. Available at: https://www.weather.co.zw (Accessed: 12 October 2024).
48. Maize and temperature change data https://www.fao.org/faostat/en/#data/QCL
49. Rainfall variability data https://climateknowledgeportal.worldbank.org/country/zimbabwe/climate-data-historical

APPENDIX A: R-Studio codes for Arima and Lstm models

APPENDIX A1: Introduction

This appendix presents the R code utilized in this study to analyze time series trends of rainfall, temperature, and maize production in Zimbabwe from 1980 to 2024. It includes steps for data preprocessing, model training, and evaluation of different analytical techniques. The section specifically focuses on two key research objectives: the first investigates the relationships between climatic variables and maize yields, while the second assesses the effectiveness of various time series analysis methods in capturing these trends. By providing detailed syntax and methodology, this appendix serves as a valuable resource for replicating the analysis and understanding the techniques applied in the study.

APPENDIX A2: R-Codes

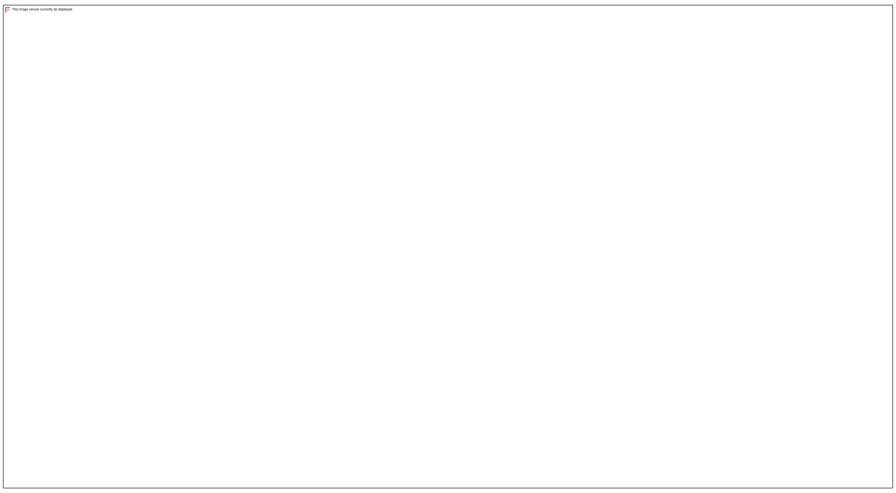
```
library(tidyverse)
library(tseries)
library(forecast)
library(zoo)
library(RSNNS)
library(Metrics)
library(ggplot2)
library(lmtest)
library(data.table)
library(psych)
library(moments)
library(corrplot)
library(dplyr)
library(urca)
library(caret)
library(car)
library(e1071)
```

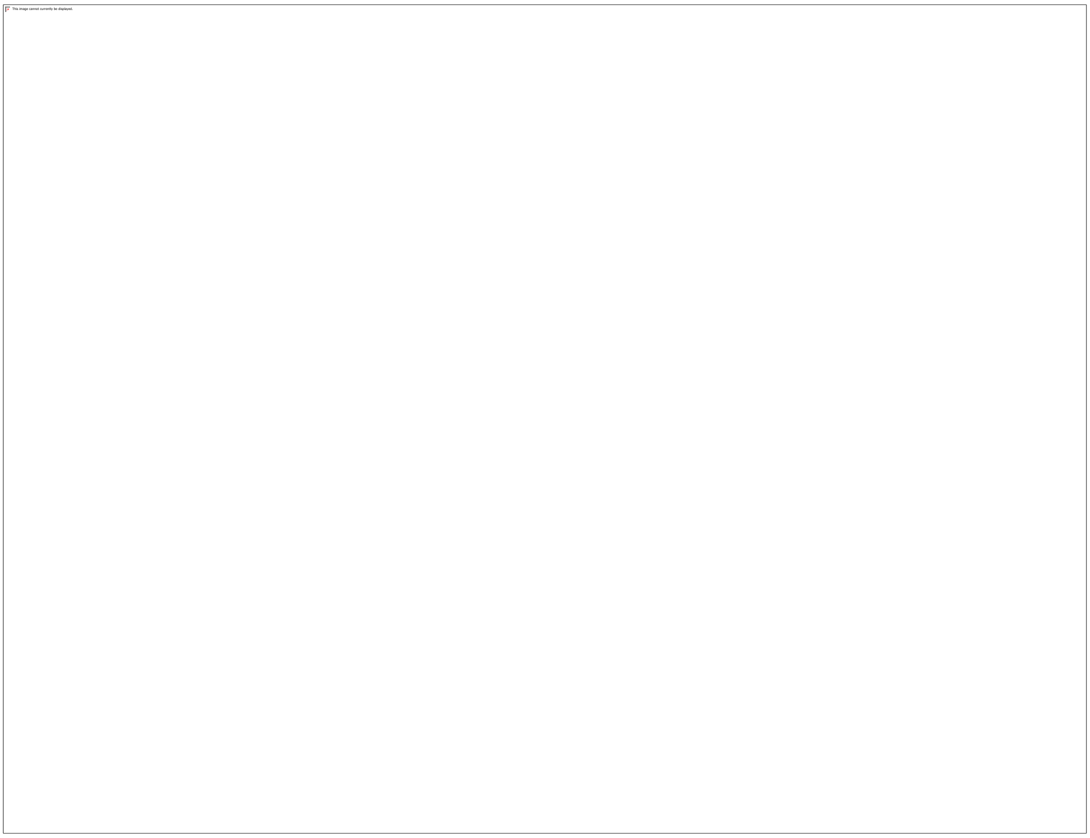
This image cannot currently be displayed.

This image cannot currently be displayed.

This image cannot currently be displayed.

```
start_year <- min(dat_stationary$Yr)
end_year <- max(dat_stationary$Yr)
n <- nrow(dat_stationary)
split_idx <- floor(0.8 * n)
train_idx <- 1:split_idx
test_idx <- (split_idx + 1):n
y_train <- dat_stationary$Ma[train_idx]
y_test  <- dat_stationary$Ma[test_idx]
xreg_train <- as.matrix(dat_stationary[train_idx, c("Ra", "Te")])
xreg_test  <- as.matrix(dat_stationary[test_idx, c("Ra", "Te")])
years_train <- dat_stationary$Yr[train_idx]
years_test  <- dat_stationary$Yr[test_idx]
par(mfrow = c(1,2))
acf(y_train, main = "ACF of Maize Production (train)")
pacf(y_train, main = "PACF of Maize Production (train)")
par(mfrow = c(1,1))
```





LSTM MODEL BUILDING

```
create_sequences <- function(data, seq_length) {
x <- list()
y <- list()
for (i in 1:(nrow(data) - seq_length)) {
x[[i]] <- as.matrix(data[i:(i + seq_length - 1), c("Ra", "Te",
"Lag1_Ra", "Lag2_Ra", "Lag1_Te", "Lag2_Te")])
y[[i]] <- data[i + seq_length, "Ma"]
}
return(list(
x = array(do.call(rbind, x), dim = c(length(x), seq_length, 6)),
y = unlist(y)
))
}
sequence_length <- 1
train_sequences <- create_sequences(train_data, sequence_length)
x_train <- train_sequences$x
y_train <- train_sequences$y
#HYPERPARAMETER TUNING
x_train_flat <- matrix(x_train, nrow = dim(x_train)[1], ncol =
dim(x_train)[2] * dim(x_train)[3])
test_sequences <- create_sequences(test_data, sequence_length)
x_test <- array(test_sequences$x, dim = c(length(test_sequences$x),
sequence_length, 6))
x_test_flat <- matrix(x_test, nrow = dim(x_test)[1], ncol =
dim(x_test)[2] * dim(x_test)[3])
set.seed(123)
lstm_model <- elman(
x_train_flat, y_train,
size = 10,  # Number of hidden units
maxit = 100,  # Maximum iterations
```

```
learnFuncParams = c(0.01),  # Learning rate
linOut = TRUE
)
```

```
> print(lstm_model)
Class: elman->rsnns
Number of inputs: 6
Number of outputs: 1
Maximal iterations: 100
Initialization function: JE_Weights
Initialization function parameters: 1 -1 0.3 1 0.5
Learning function: JE_BP
Learning function parameters: 0.01
Update function:JE_Order
Update function parameters: 0
Patterns are shuffled internally: FALSE
Compute error in every iteration: TRUE
Architecture Parameters:
$size
[1] 10

All members of model:
 [1] "nInputs"               "maxit"                 "initFunc"
 [4] "initFuncParams"        "learnFunc"             "learnFuncParams"
 [7] "updateFunc"            "updateFuncParams"      "shufflePatterns"
[10] "computeIterativeError" "snnsObject"            "archParams"
[13] "IterativeFitError"     "fitted.values"         "nOutputs"
```

```
# 16. LSTM Prediction & Performance
predictions <- predict(lstm_model, x_test_flat)
actual_values <- test_data$Ma[(sequence_length + 1):nrow(test_data)]
mse_value <- mse(actual_values, predictions)
rmse_value <- rmse(actual_values, predictions)
mae_value <- mae(actual_values, predictions)
mape_value <- mean(abs((actual_values - predictions) / actual_values))
* 100
rsquared_value <- 1 - sum((actual_values - predictions)^2) /
sum((actual_values - mean(actual_values))^2)
cat("\nLSTM Model Performance Metrics:\n")
cat("MSE:", mse_value, "\n")
cat("RMSE:", rmse_value, "\n")
cat("MAE:", mae_value, "\n")
cat("MAPE:", mape_value, "%\n")
cat("R-squared:", rsquared_value, "\n")
# 17. LSTM Actual vs Predicted Plot
comparison_df <- data.frame(
Year = seq(1980 + train_size + sequence_length, 1980 + nrow(DATART) -
1),
Actual = actual_values,
Predicted = predictions
)
ggplot(comparison_df, aes(x = Year)) +
geom_line(aes(y = Actual, color = "Actual"), size = 1.5) +
geom_line(aes(y = Predicted, color = "Predicted"), size = 1.5, linetype
= "dashed") +
labs(title = "Actual vs Predicted Maize Production (LSTM)",
y = "Maize Production (normalized)", x = "Year") +
scale_color_manual(values = c("Actual" = "red", "Predicted" = "blue"))
+
```

```
theme_minimal() +
theme(text = element_text(size = 14, face = "bold"),
legend.title = element_blank()) +
guides(color = guide_legend(override.aes = list(size = 2)))
library(ggplot2)
# Create the data frame for actual and predicted values
comparison_df <- data.frame(
Year = c(2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024),
Actual = c(0.47471, 0.48587, 0.1671, 0.27125, 0.3408, 0.4492, 0.2667,
0.11062, 0.4615),
Predicted = c(0.56574, 0.51923, 0.18968, 0.29864, 0.36314, 0.38863,
0.24864, 0.18937, 0.28973)
)
# Convert Year to a date format for better plotting
comparison_df$Date <- as.Date(paste(comparison_df$Year, "-01-01",
sep=""))
# Smoothed plot
ggplot(comparison_df, aes(x = Date)) +
geom_line(aes(y = Actual, color = "Actual"), size = 1.5, linetype =
"solid") +
geom_line(aes(y = Predicted, color = "Predicted"), size = 1.5, linetype
= "dashed") +
labs(title = "Actual vs Predicted Maize Production",
y = "Maize Production (normalized)",
x = "Date",
caption = "LSTM Model: Hidden Neurons = 10, Learning Rate = 0.01,
Epochs = 100") +
scale_color_manual(values = c("Actual" = "red", "Predicted" = "blue"))
+
theme_minimal() +
theme(text = element_text(size = 14, face = "bold"),
legend.title = element_blank()) +
guides(color = guide_legend(override.aes = list(size = 2))) +  #
Increase legend line size
scale_x_date(date_labels = "%Y")  # Format x-axis to show years
if (!require(Metrics)) install.packages("Metrics")
library(Metrics)
# Calculate actual values for comparison
actual_values <- comparison_df$Actual
predicted_values <- comparison_df$Predicted
# Performance metrics
mse_value <- mse(actual_values, predicted_values)
rmse_value <- rmse(actual_values, predicted_values)
mae_value <- mae(actual_values, predicted_values)
mape_value <- mean(abs((actual_values - predicted_values) /
actual_values)) * 100
rsquared_value <- 1 - sum((actual_values - predicted_values)^2) /
sum((actual_values - mean(actual_values))^2)
# Print performance metrics
cat("Performance Metrics:\n")
cat("MSE:", mse_value, "\n")
cat("RMSE:", rmse_value, "\n")
cat("MAE:", mae_value, "\n")
cat("MAPE:", mape_value, "%\n")
cat("R-squared:", rsquared_value, "\n")
```

LSTM FORECASTING FOR NEXT 5YEARS

This image cannot currently be displayed.

APPENDIX B: TRANSFORMED DATA .

APPENDIX B1: TRAINED DATA 80%

Yr	Ma	Ra	Te	Lag1_Ra	Lag2_Ra	Lag1_Te	Lag2_Te	RollingMean_	RollingSD_Ra	RollingMean_	RollingSD_Te
1982	0.58654394	0.06602972	0.16786141	0.82057407	0.48769281	0	0.1916264	0.4723194	0.6939311	0	0.21849118
1983	0.22216698	0.10191895	0.79271207	0.06602972	0.82057407	0.2651637	0	0.2744807	0.786883	0.3606707	1
1984	0.31260006	0.28535282	0.35603345	0.10191895	0.06602972	0.8266237	0.2651637	0	0.1843023	0.6073171	0.63268625
1985	1	0.71889802	0.09259259	0.28535282	0.10191895	0.4342458	0.8266237	0.3348143	0.573976	0.5689024	0.7140901
1986	0.88548721	0.55653617	0.08363202	0.71889802	0.28535282	0.1975309	0.4342458	0.5679584	0.3826658	0.2070122	0.22439896
1987	0.31180406	0.11171612	0.70310633	0.55653617	0.71889802	0.1894793	0.1975309	0.4789113	0.5691322	0.3841463	0.71773788
1988	0.80260695	0.49095178	0.34468339	0.11171612	0.55653617	0.7461084	0.1894793	0.3620122	0.4239389	0.5128049	0.60918084
1989	0.67176432	0.45745112	0.20131422	0.49095178	0.11171612	0.4240472	0.7461084	0.3111978	0.3648889	0.5728659	0.47978601
1990	0.65270751	0.44581191	0.44623656	0.45745112	0.49095178	0.2952228	0.4240472	0.4825342	0	0.4417683	0.14635324
1991	0.49627347	0.20148982	0.28673835	0.44581191	0.45745112	0.5152979	0.2952228	0.3340877	0.2369165	0.4121951	0.14945439
1992	0	0.06702158	0.75149343	0.20148982	0.44581191	0.3719807	0.5152979	0.1338614	0.3298339	0.6929878	0.4248585
1993	0.68979482	0.51866321	0.35364397	0.06702158	0.20148982	0.7895867	0.3719807	0.1712221	0.407827	0.6457317	0.46202456
1994	0.70856129	0.08568479	0.15113501	0.51866321	0.06702158	0.4320988	0.7895867	0.1118332	0.4540752	0.5765244	0.5954401
1995	0.21210089	0.06064532	0.58542413	0.08568479	0.51866321	0.2501342	0.4320988	0.1085632	0.4579385	0.4917683	0.3784481
1996	0.69074531	0.58997611	0	0.06064532	0.08568479	0.640365	0.2501342	0.145135	0.5384045	0.3112805	0.59169931
1997	0.4828689	0.79381402	0.24432497	0.58997611	0.06064532	0.1143317	0.640365	0.5082891	0.6945045	0.3588415	0.5673987
1998	0.33819756	0.51319785	0.74133811	0.79381402	0.58997611	0.3338701	0.1143317	0.7403743	0.2378683	0.4384146	0.77359361
1999	0.50471919	0.55244727	0.27718041	0.51319785	0.79381402	0.7804616	0.3338701	0.7211282	0.2514316	0.579878	0.52777776
2000	0.51001622	0.94364601	0.01135006	0.55244727	0.51319785	0.3633924	0.7804616	0.7979674	0.4197632	0.4609756	0.75310272
2001	0.47217388	0.90714951	0.24133811	0.94364601	0.55244727	0.1245303	0.3633924	1	0.3769101	0.2057927	0.19854046
2002	0.09847857	0	0.49283154	0.90714951	0.94364601	0.3311863	0.1245303	0.7166852	1	0.3158537	0.43634452
2003	0.28258627	0.3459779	0.42771804	0	0.90714951	0.5571659	0.3311863	0.4101795	0.8497887	0.5283537	0.16475713
2004	0.80907181	0.67622768	0.18339307	0.3459779	0	0.4986581	0.5571659	0.2917545	0.6156821	0.4987805	0.24505904
2005	0.36248412	0.32638355	0.85663082	0.67622768	0.3459779	0.2791197	0.4986581	0.4591357	0.3387133	0.6844512	0.68256216
2006	0.66319411	0.62570341	0.0734767	0.32638355	0.67622768	0.884058	0.2791197	0.602589	0.3240839	0.5036585	0.8874138
2007	0.46523255	0.52999879	0.33154122	0.62570341	0.32638355	0.1803543	0.884058	0.5275976	0.2532148	0.5792683	0.82607715
2008	0.05437736	0.30431966	0.28315412	0.52999879	0.62570341	0.4122383	0.1803543	0.5162824	0.2769778	0.2865854	0.18112562
2009	0.13709906	0.47323995	0.07885305	0.30431966	0.52999879	0.3687601	0.4122383	0.4380937	0.183814	0.2893293	0.17360471
2010	0.33676615	0.55163759	0.71863799	0.47323995	0.30431966	0.1851852	0.3687601	0.4491908	0.2014151	0.4868902	0.64801279
2011	0.26299542	0.68147039	0.38172043	0.55163759	0.47323995	0.7600644	0.1851852	0.6426073	0.1598988	0.5371951	0.63142518
2012	0.29765772	0.28128416	0.44563919	0.68147039	0.55163759	0.4573269	0.7600644	0.5441654	0.3535681	0.7243902	0.28397894
2013	0.23412769	0.43241164	0.31839904	0.28128416	0.68147039	0.5147611	0.4573269	0.4830221	0.3494926	0.520122	0
2014	0.24859333	0.35937816	0.27837515	0.43241164	0.28128416	0.4004294	0.5147611	0.3178416	0.1020087	0.467378	0.05840718
2015	0.11390163	0.0133193	0.80824373	0.35937816	0.43241164	0.3644659	0.4004294	0.1804196	0.3921371	0.652439	0.56986891

APPENDIX B2: TESTED DATA 20%

Yr	Ma	Ra	Te	Lag1_Ra	Lag2_Ra	Lag1_Te	Lag2_Te	RollingMean_	RollingSD_Ra	RollingMean_	RollingSD_Te
2016	0.06079072	0.2329865	0.9856631	0.0133193	0.3593782	0.8405797	0.3644659	0.07814722	0.2967329	0.9929878	0.7494875
2017	0.47470581	1	0.2921147	0.2329865	0.0133193	1	0.8405797	0.40668113	0.9675935	1	0.7305427
2018	0.48586838	0.3810979	0.4904421	1	0.2329865	0.3768116	1	0.59529124	0.7501448	0.8378049	0.7229414
2019	0.16709622	0.5771426	0.827957	0.3810979	1	0.5550188	0.3768116	0.77178686	0.5729507	0.7573171	0.5104552
2020	0.34080005	0.4800008	0.4898447	0.5771426	0.3810979	0.8582931	0.5550188	0.50511258	0.1459207	0.8582317	0.323605
2021	0.44919671	0.3794178	0.2293907	0.4800008	0.5771426	0.554482	0.8582931	0.50425097	0.1475713	0.725	0.5823689
2022	0.26669798	0.5629732	0.3853047	0.3794178	0.4800008	0.3204509	0.554482	0.49698436	0.1339765	0.4990854	0.1660873
2023	0.1106196	0.4510749	0.692951	0.5629732	0.3794178	0.4605475	0.3204509	0.48215009	0.1351339	0.6027439	0.4241846
2024	0.46149791	0.4976317	1	0.4510749	0.5629732	0.7369834	0.4605475	0.54277439	0.0641209	0.9960366	0.6001628