Modelling the Effect of Weather Variables on Maize Production using Copula Methods

By

Jefferson Mwangi Chege

138063

Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Mathematical Finance at Strathmore University

Institute of Mathematical Sciences
Strathmore University
Nairobi, Kenya

June, 2025

Declaration and Approval

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the dissertation contains no material previously published or written by another person except where due reference is made in the dissertation itself.

Name: **Jefferson Mwangi Chege**
Signature:
Date:

Approval

The dissertation of Jefferson Mwangi Chege was reviewed and approved for examination by the following:

Dr. Stanley Sewe
Lecturer,
Institute of Mathematical Sciences, Strathmore University.

Dr. Godfrey Madigu
Dean,
Institute of Mathematical Sciences, Strathmore University.

Prof. Bernard Shibwabo
Director,
Office of Graduate Studies, Strathmore University.

Abstract

This study investigates the interplay between weather variables and maize production, employing advanced statistical and copula modeling techniques. The framework integrates: data pre-treatment (autocorrelation, stationarity testing and detrending); exploratory dependence analysis using Pearson correlation coefficient, Kendall's tau, and Spearman's rho, supplemented by graphical diagnostics; bivariate and trivariate copula selection using semi-parametric Canonical Maximum Likelihood estimation and further modeling by vine copulas; and goodness-of-fit validation by AIC and BIC criteria. This structured approach ensures robust modeling of complex dependencies between climatic and agricultural variables while addressing temporal biases and non-linear interactions. Maize production, precipitation and temperature data from Uasin Gishu County, Kenya over the period 1990 to 2020 was used in the analysis. This study establishes that copula-based models, particularly when enhanced with semi-parametric methods and vine structures, significantly improve the modeling of nonlinear, multivariate dependencies between weather variables and maize production compared to traditional linear approaches.Particularly, copula models overcome linear correlation limitations and estimate the maize production-weather variables nexus with greater precision. The semi-parametric approach also ensures robustness through rank-based margins. Vine structures identified conditional relationships, with precipitation-temperature conditioning amplifying maize-precipitation dependence. Nonparametric vine copulas used for comparison with the parametric class offered superior fit, though interpretability favored parametric models. Hence, this study demonstrated copulas' utility in disentangling multivariate dependencies in agro-climatic systems, advocating for nuanced, data-driven approaches in agricultural planning.

Keywords: Maize production, Copula modeling, Weather variables, Nonlinear dependencies, Semi-parametric estimation

Table of Contents

List of Figures

List of Tables

Acknowledgement

First and foremost, I would like to express my deepest gratitude to God Almighty for His guidance, strength, and countless blessings throughout this journey. Without His grace, none of this would have been possible.

I am profoundly grateful to the DAAD (German Academic Exchange Service) for awarding me the scholarship that made my studies and research possible. Your support has been instrumental in shaping my academic and professional path, and I am truly honored to have been a beneficiary of your generosity.

To my supervisor, Dr. Stanley Sewe, I extend my heartfelt thanks for your unwavering support, insightful guidance, and patience throughout this process. Your expertise, encouragement, and constructive feedback have been invaluable in shaping this thesis and my growth as a researcher. I am deeply appreciative of the time and effort you dedicated to helping me succeed.

To my beloved parents, I owe an immeasurable debt of gratitude. Your unconditional love, sacrifices, and constant encouragement have been my foundation. Thank you for believing in me and for always pushing me to strive for excellence. This achievement is as much yours as it is mine.

To my classmates and friends, thank you for the camaraderie, shared experiences, and moral support throughout this journey. Your encouragement during challenging times and the memories we created together have made this experience truly unforgettable.

Finally, I would like to acknowledge all those who, directly or indirectly, contributed to the completion of this thesis. Your support, whether big or small, has meant the world to me.

Thank you all.

Dedication

This dissertation is dedicated to Favor Wangechi, our baby sister in whom I delight in.

Chapter 1: Introduction

1.1 Background to the study

Maize (Zea mays L.) is a cornerstone of global food security, serving as a staple crop for over 1.2 billion people in sub-Saharan Africa and a critical source of calories, livelihoods, and economic activity FAO (2021). In Kenya, maize accounts for 40% of caloric intake and 30% of agricultural GDP, with smallholder farmers contributing over 75% of national production. However, climate change poses an existential threat to maize-dependent economies. Rising temperatures, erratic rainfall, and extreme weather events—such as droughts and floods—have disrupted traditional growing seasons, reduced yields, and intensified food insecurity. Between 1980 and 2020, maize yields in East Africa stagnated at 1.5–2.0 tons/ha, far below the global average of 5.6 tons/ha, with climate variability accounting for 32–39% of yield losses AGRA (2021).

Uasin Gishu County, Kenya's primary maize-producing region, exemplifies these challenges. It is located in the former Rift Valley province in Kenya with a size of approximately $3,350km^2$. It is classified in three ecological zones namely: Upper highlands, Lower highlands and Upper midlands. Dubbed the "breadbasket of Kenya," the county contributes 25% of the nation's maize output, sustaining over 1 million livelihoods. Yet, its rainfed agricultural system is acutely vulnerable to climatic shifts. Historical data reveals a 1.2°C temperature rise since 1990 and a 15% decline in long-term precipitation, exacerbating water stress during critical growth stages. The 2010–2011 drought alone caused a 40% drop in maize production, triggering national food shortages and price spikes. Such volatility underscores the urgent need for robust models to quantify climate-agriculture linkages and inform adaptive strategies.

Traditional approaches to studying weather-crop relationships, such as linear regression or Pearson correlation, often fail to capture nonlinear dependencies, tail risks (e.g., simultaneous extremes), and temporal biases inherent in agro-climatic systems. For instance, while maize yield correlates positively with moderate rainfall, excessive precipitation during flowering can induce waterlogging and disease, a nonlinear relationship obscured by conventional methods Lobell and Burke (2010). Similarly, temperature and rainfall often exhibit inverse dependencies (e.g., drought-heatwave coupling), which linear models inadequately represent.

Copula theory, a statistical framework for modeling multivariate dependencies, addresses these limitations by decoupling marginal distributions from their joint behavior. Originally applied in finance and hydrology, copulas are increasingly recognized for their utility in agriculture. They enable the analysis of tail dependencies (e.g., joint extremes of high temperature and low rainfall) and conditional relationships (e.g., yield responses to precipitation under specific temperature thresholds). Recent studies in India and Brazil demonstrate copulas' superiority in predicting crop failures under compound climate stresses, outperforming traditional models by 20–30% (Karthikeyan et al. (2020); Ribeiro et al. (2019)). However, their application in sub-Saharan Africa remains sparse, particularly in smallholder contexts where data scarcity and non-stationarity complicate modeling.

This study bridges this gap by employing a semi-parametric copula framework—combining non-parametric margins (to handle skewed climatic data) with parametric copulas (to model dependencies). This approach is particularly suited to Uasin Gishu's context, where limited historical data and shifting climate baselines challenge fully parametric models. By integrating vine copulas (for high-dimensional dependencies) and goodness-of-fit tests (AIC/BIC), the methodology advances localized risk assessment, offering actionable insights for farmers, policymakers, and insurers.

Kenya's Climate-Smart Agriculture Strategy (2020–2030) prioritizes climate-resilient maize systems, yet implementation is hindered by insufficient granularity in risk models. Concurrently, global initiatives like the UNFCCC's Nairobi Work Programme emphasize localized adaptation planning. This study aligns with these goals, providing empirical tools to optimize irrigation investments, crop insurance design, and early-warning systems. In summary, the convergence of climate urgency, methodological innovation, and policy demand positions this research as a critical contribution to sustainable agricultural planning in maize-dependent regions.

1.2 Statement of the problem

Maize production in Kenya is vital for food security and rural economies, yet climate variability increasingly threatens its stability. Current analytical methods, like linear regression and Pearson correlation, struggle to capture the complex, nonlinear relationships between weather patterns and crop yields, especially in smallholder farming systems. These traditional approaches oversimplify climate impacts by assuming steady relationships, missing critical risks that cause significant yield losses. While studies in other regions show that extreme temperatures and compounding climate stressors are underestimated by these methods, similar analyses in sub-Saharan Africa remain limited. Existing models also fail to address skewed climate data and shifting weather patterns due to climate change, leaving policies and insurance programs misaligned with on-ground risks.

To address these gaps, advanced statistical tools like copula methods—proven in fields like hydrology and finance—could offer better insights. Though used successfully in India and Brazil to predict crop failures during extreme weather, copula applications in Africa are rare and often restricted to basic two-variable models. Current approaches also rely on rigid assumptions that don't fit Africa's variable climate data. This study proposes a flexible, semi-parametric framework combining non-parametric data adjustments with copula models to analyze three-way interactions (maize yields, temperature, and rainfall). By incorporating methods to correct time-related biases and extend beyond simplistic models, the research aims to provide practical tools for adaptive farming, improved policies, and fairer insurance programs tailored to Kenya's climate challenges.

1.3 Research objectives

The **main objective** was to model the joint and interactive effects between maize production, precipitation and temperature using copulas.

Specific objectives were:

1. To estimate the dependence structures between maize production, precipitation and temperature using bivariate copulas.
2. To estimate the joint dependence structures between maize production, precipitation and temperature using trivariate copulas.
3. To model the conditional dependence between maize production, precipitation and temperature using vine copulas.

1.4 Research hypotheses

The hypotheses for this study was as follows:

- Null hypothesis (H0): Traditional linear models (e.g., Pearson correlation) adequately explain maize production variability.
- Alternative hypothesis (H1): The proposed semi-parametric copula framework better captures nonlinear dependencies between climate variables (temperature, precipitation) and maize production.

1.5 Scope of the study

Due to time and resource constraints:

- This study was limited to Uasin Gishu county, one of the 47 counties in the Republic of Kenya.
- This study focused on only two variables within the category of weather factors which have an influence maize production i.e temperature and precipitation. For the purpose of this study, the effect of other exogenous factors such as soil structure, farm input, farm size, available technology was assumed to be negligible.

1.6 Significance of the study

Farmers: Copula methods capture non-linear dependencies between weather variables, enabling farmers to anticipate compound risks like droughts coinciding with heatwaves. This informs adaptive strategies such as crop diversification, irrigation scheduling, or planting drought-resistant varieties. Copulas also model extreme weather events (e.g., heavy rainfall or prolonged dry spells) more accurately than traditional correlation methods. Farmers can use these insights to prepare for rare but catastrophic events. By understanding how weather variables jointly affect yields, farmers can optimize input use (e.g., fertilizer, water) under specific climatic conditions.

Policymakers: Copula-based models reveal how shifts in temperature and precipitation patterns interact to impact agriculture. This supports evidence-based policies for climate adaptation (e.g., subsidizing resilient crops or irrigation infrastructure). Policymakers can quantify the likelihood of simultaneous crop failures across regions due to correlated weather shocks. This informs strategic grain reserves or import/export regulations during crises. Moreover, identifying regions most vulnerable to compound weather risks e.g., high temperature and/or low rainfall helps prioritize investments in rural infrastructure, early-warning systems, or insurance schemes.

Risk managers: Copulas model tail dependencies between weather variables, allowing insurers to design better-indexed insurance products (e.g., parametric insurance) that reflect true joint risks of crop failure. In addition, financial institutions can assess spatial dependencies in weather risks (e.g., droughts affecting multiple regions) to avoid overexposure in agricultural loans or investments. Hence, copulas enable scenario analysis for compound extremes such as heatwaves and floods, helping risk managers evaluate systemic threats to agribusiness supply chains.

Chapter 2: Literature Review

2.1 Introduction

This chapter provides a summary of the works done in the subject of study. Prior studies have investigated the effect of climatic variables on maize yields using other methodologies such as production function models, econometric methods and copulas. A thorough literature review was conducted and clearly highlighted the research gap this study sought to fill.

2.2 Modeling the effect of weather variables on maize yields using econometric models

Recent studies have employed production function models to quantify the relationship between climatic variables and agricultural productivity. Dwamena et al. (2022) conducted a seminal analysis in Ghana's Ashanti region, examining temperature, rainfall, and relative humidity effects on maize, cassava, and yam yields from 1990 to 2020. Their multiple linear regression and correlation analyses revealed that 44–57% of maize yield variability was attributable to fluctuations in minimum and maximum temperatures. Notably, Pearson's correlation coefficients demonstrated significant inverse relationships: minimum and maximum temperatures reduced maize yields by 66% (r = -0.66) and 75% (r = -0.75), respectively. The derived regression equation indicated that a 1% increase in minimum and maximum temperatures reduced maize yields by 64 kg/ha and 34 kg/ha, respectively. Cumulatively, 64% of maize yield variations were explained by these climatic factors, underscoring temperature as the dominant stressor in tropical agroecosystems.

Complementing these findings, Kariuki et al. (2018) employed a Cobb-Douglas production function and OLS regression to analyze climate-maize yield linkages in Kenya (1970–2014), uncovering nonlinear dynamics. Rainfall during long rains

(March–May) and short rains (October–December) exhibited quadratic effects: initial increases enhanced yields, but excessive rainfall reduced productivity. Conversely, January–February rainfall negatively impacted yields, as dry conditions during this non-growing period facilitated harvesting and land preparation. Temperature displayed an inverted U-shaped relationship, peaking at 18.25°C; a 1°C rise beyond this threshold decreased yields by 0.33 tonnes/ha. Temperature variability weakly diminished yields, aligning with Cabas et al. (2010) and Blanc (2011) but contrasting Akpalu et al. (2008), highlighting regional disparities. The study reinforces Dwamena et al. (2022), emphasizing diminishing marginal returns and threshold effects of climatic extremes on agricultural outcomes.

Similarly, Riaman et al. (2022) applied a Cobb-Douglas framework to evaluate weather-related risks for Indonesian rice farmers (2008–2021). Using Value-at-Risk (VaR), Expected Shortfall (ES), and Tail-VaR metrics, they demonstrated that precipitation and temperature increases initially boosted yields but became detrimental beyond critical thresholds, reinforcing the non-linear climate-yield relationships posited by Kariuki et al. (2018). However the authors admitted that limited historical loss data and distribution assumptions on normality of the data may likely understate tail risks.

2.3 Modeling the effect of weather variables on maize yields using Ricardian models

Ricardian models, which link net farm revenues to climatic variables, further elucidate these dynamics. Kimani (2017) conducted a study analyzing the impact of climate on agriculture with a focus on maize production in Uasin Gishu and Trans-Nzoia counties in Kenya. The study employed the Ricardian model to investigate the relationship between net farm revenue and climate variables. The results indicated that an increase in precipitation during the formative period of maize had a positive effect on net farm revenue per acre while an increase in temperature during the same period had a negative effect on net farm revenue per acre. Another notable result was the fact that net farm revenue tended to be more responsive to variations in temperature compared to changes in rainfall hence the conclusion that changes in temperature is a significant variable and its effect far-reaching in comparison to changes in precipitation.

Similarly, Bello and Maman (2015) examined the impact of climate variables on crop revenue in Niger's Dosso and Maradi regions using the Ricardian model, integrating primary household surveys and secondary climate data. Key findings reveal that rising temperatures and declining precipitation significantly reduce crop yields, with a 1°C temperature increase and 10% precipitation decrease leading to a 26% yield decline. Nonlinear relationships were identified: wet-period temperature boosts revenue up to a threshold before becoming detrimental, while dry-period heat consistently harms yields. Similarly, wet-season rainfall increases revenue until excessive amounts trigger flooding. Socioeconomic factors—input costs, labor shortages, and input availability—critically influence farmer resilience. Farmers perceive climate shifts and adapt via irrigation (26%, dominant in Dosso), livestock diversification, crop mixing, and tree planting. Statistical models (significant at 1% via S. Fisher test) explain 44% of net revenue variance, aligning with literature on nonlinear climate-revenue dynamics.

2.4 Modeling the effect of weather variables on maize yields using crop simulation models

Crop simulation models, such as GLAM-Maize and EcoCrop, enable spatially explicit projections of climate change impacts. Chisanga et al. (2022) reviewed 27 such models, identifying temperature variability as the primary driver of global maize yield uncertainty, particularly in sub-Saharan Africa. Their meta-analysis emphasized calibration challenges, as improper parameterization inflated prediction errors by up to 30%.

Li et al. (2019b) compared U.S. maize yield losses under extreme drought and excessive rainfall (1981–2016), finding both stressors reduced yields by 18–22% relative to long-term trends. Soil texture mediated these effects; clay-rich soils exacerbated waterlogging losses, whereas sandy soils amplified drought impacts.

Jones and Thornton (2003) projected a 10% decline in African and Latin American maize production by 2055 using process-based simulations. Regionally, yields increased by 8% in Ethiopia due to elevated CO2 fertilization, but collapsed by 94% in Venezuela under high-emission scenarios. These disparities underscore the need for context-specific adaptation strategies, such as heat-tolerant germplasm in vulnerable regions.

2.5 Modeling the effect of weather variables on maize yields using copula models

On the other hand, copula methods, which model multivariate dependencies without assuming normality, have gained traction in climate-risk analyses. Ribeiro et al. (2019) investigated the application of copula methods to assess drought risk in rainfed wheat and barley systems in the Iberian Peninsula (IP), focusing on the joint probability of crop yield anomalies and drought hazards for the period 1986 to 2012. Bivariate Archimedean copulas (e.g., Clayton, Gumbel) were used to model dependencies between yield anomalies and drought indices Standardized Precipitation Evapotranspiration Index (SPEI) and Vegetation Condition Index (VCI), with these indices employed to capture multi-scalar drought impacts. The results showed that extreme drought events (SPEI $<$ -1.5) disproportionately increased crop loss risks, with barley in cluster 1 and wheat in cluster 2 showing higher vulnerability. Similarly, Conditional Probability of Non-Exceedance (CPNE) revealed a 14.1% risk of wheat loss in cluster 2 during droughts, compared to 3.97% in non-drought conditions. The study focused on yield anomalies rather than drought characteristics alone, hence offering actionable insights for farmers. Equally, the results demonstrated asymmetric dependencies between extreme droughts and affect yields, a key relationship which is overlooked by linear models.

Recently, Unnikrishnan et al. (2024) employed multivariate C-vine copula analysis to quantify the joint dependence structure between temperature anomalies and strawberry yield in Santa Maria, California (2011–2019). By modeling nonlinear dependencies, the analysis revealed that temperature anomalies exceeding 3°F significantly elevated the conditional probability of extreme yield losses—a relationship obscured by univariate methods. The copula framework demonstrated that yield losses are not merely proportional to temperature shifts but are shaped by threshold effects and nonlinear interactions, underscoring the need for probabilistic risk assessment in climate-agriculture models. Nevertheless, the findings were regionally constrained to Santa Maria's strawberry systems with the authors pointing out caution when extrapolating to rainfed or other crops (e.g., maize). Moreover, the omission of pest and disease impacts limit applicability to non-commercial, organic systems. Lastly, the short data span of 9 years may not have fully captured long-term climate variability as a timeframe of at least 30 years is considered ideal.

2.6 Research gap

Linear regression and correlation coefficients (Pearson, Spearman, Kendall) are foundational tools in statistical analysis due to their simplicity. Pearson's correlation measures linear relationships, while Spearman's and Kendall's rank-based coefficients assess monotonic associations. These methods are computationally efficient, require minimal assumptions for basic applications, and yield intuitive results. For instance, Lobell and Burke (2010) effectively used linear regression to isolate the impact of temperature on crop yields in sub-Saharan Africa, demonstrating their utility in identifying broad trends. In fact, linear models underpin many agricultural risk frameworks. The FAO's Global Information and Early Warning System (GIEWS) relies on linear correlations between rainfall anomalies and yield deviations to trigger food security alerts FAO (2021). Similarly, Kenya's National Drought Management Authority uses Spearman's rho to link precipitation deficits to pastoralist losses, enabling rapid humanitarian responses. Moreover, linear methods perform adequately in data-scarce contexts. In a study of Ethiopian maize systems, Di Falco et al. (2011) found that linear regression provided "sufficiently accurate" risk assessments for policymakers, despite ignoring nonlinearities.

Unfortunately, linear models assume proportionality between variables, which collapses under extreme conditions. Schlenker and Roberts (2009) demonstrated that maize yields decline nonlinearly above temperature thresholds (e.g., 30°C), rendering Pearson's correlation inadequate for quantifying heatwave impacts. Similarly, Karthikeyan et al. (2020) showed that linear models underestimate yield losses during concurrent droughts and heatwaves by 20–30% in India. As a matter of fact, climate change invalidates the stationarity assumption implicit in linear models. In Uasin Gishu County, a 1.2°C temperature rise since 1990 has shifted yield-response curves, making historical correlations obsolete. Tschora and Cherubini (2020) critiqued linear approaches as "statistical relics" in non-stationary climates. Most notably, linear frameworks analyze variables in isolation, neglecting compounding risks. For example, while moderate rainfall correlates positively with yields, excessive rainfall during flowering interacts with temperature to amplify fungal outbreaks—a relationship invisible to Pearson's rho Ribeiro et al. (2019).

The deficiencies of linear models elevates the utility of copula based methods. Copulas separate marginal distributions from dependence structures, enabling analysis of nonlinear and tail dependencies. In Brazil, vine copulas improved soybean failure predictions by 35% compared to linear models by capturing heat-rainfall interactions Ribeiro et al. (2019). Similarly, Aas et al. (2009) demonstrated that copulas outperform multivariate regression in hydrology by modeling flood-drought cycles as joint extremes. Specifically, semi-parametric copulas, which use empirical margins, adapt to shifting climate baselines.Chen et al. (2006) showed that semi-parametric frameworks reduce prediction errors by 15–20% in non-stationary contexts compared to fully parametric models.

While linear methods retain utility for preliminary analyses in data-scarce settings, copulas offer unparalleled advantages in modeling agro-climatic systems. They quantify risks from compound extremes (e.g., drought-heatwaves), which linear models overlook. In particular, vine copulas unravel conditional relationships (e.g., yield responses to rainfall under specific temperature thresholds), critical for adaptive farming. Lastly, the semi-parametric methodology bypasses distributional assumptions, aligning with skewed, non-stationary climatic data.

2.7 Conceptual framework

By integrating these steps, the study advanced beyond linear paradigms, offering a replicable framework for climate-resilient agricultural planning in data-constrained regions.

Data Pre-Treatment: Autocorrelation and Stationarity testing using parametric tests and visualized plots addressed any temporal biases while detrending removed secular trends from maize production data.

Exploratory Dependence Analysis: Linear Metrics such as Pearson, Spearman, and Kendall coefficients identified baseline associations while graphical diagnostics like scatterplots and QQ-plots visualized nonlinearities.

Copula Modeling: Bivariate/Trivariate Copulas of the families Gaussian, Student-t, Clayton, Gumbel, and Frank were tested via semi-parametric Canonical Maximum Likelihood (CML). Further, vine copulas decomposed trivariate dependencies (maize × temperature × precipitation) into pairwise conditional copulas.

Model Validation: Goodness-of-Fit tests such as Akaike/Bayesian Information Criteria (AIC/BIC) selected optimal copulas while nonparametric copula families were compared with parametric copulas when modeling vine copulas to highlight any parametric bias.

Chapter 3: Methodology

3.1 Introduction

This chapter outlines the general methodology used to conduct the study. It highlights the research design, sampling design, data collection method and data analysis techniques of the study.

3.2 Research Design

This study employs a quantitative research design, which emphasizes the systematic collection and analysis of empirical data to investigate relationships between variables through statistical and mathematical modeling. The design aligns with the objective of rigorously testing theoretical relationships, evaluating causal linkages, and quantifying dependencies among climatic and agricultural variables. By applying statistical techniques—including correlation analysis and copula-based dependence structures—the study seeks to derive objective insights into how weather variables (temperature, precipitation) influence maize production in Uasin Gishu County. Through this approach, the research transcends descriptive analysis, offering predictive tools to inform climate-resilient agricultural practices.

3.3 Data Description

The study focused on Uasin Gishu County, located in the former Rift Valley Province of Kenya. Secondary data covering the period from 1990 to 2020 were utilized in this study.The maize production dataset, which includes data on the area under production (in hectares), annual maize production (in tonnes), and annual maize yield (in tonnes per hectare), was sourced from the Kenya National Bureau of Statistics. The weather variables dataset containing monthly precipitation (in mm) and temperature (in degrees celsius) was retrieved from the FAO AQUASTAT online portal (https://aquastat.fao.org/climate-information-tool/).

3.4 Model Specification

3.4.1 Correlation and rank-based methods

The correlation coefficient is a widely recognized statistical measure employed to quantify the strength and direction of linear associations between variables. Rooted in probability theory, it serves as a foundational tool in empirical research for evaluating bivariate dependencies, particularly in contexts where parametric assumptions of normality and homoscedasticity hold.

For two random variables X and Y, the Pearson's correlation coefficient is expressed as follows:

$$\rho(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{3.1}$$

where X_i & Y_i refer to the individual observations from the datasets,$\bar{X}$ & $\bar{Y}$ refer to the mean for both X & Y and n refers to the number of observations.

While its computational simplicity and interpretability render it prevalent in fields ranging from econometrics to climatology, its utility is constrained by an inherent limitation: the presumption of linearity, which obscures nonlinear and tail-dependent relationships. Nevertheless, its integration into exploratory data analysis remains indispensable for preliminary hypothesis formulation and model diagnostics.

Next, this study employed the use of non-parametric measures of association such as Kendal's Tau and Spearman's rho for further exploratory dependence analysis.

Definition 1 : Let (x_i, y_i) and (x_j, y_j) be two observations of a vector (X.Y) of continuous random variables. X and Y are said to be concordant if $x_i < x_j$ and $y_i < y_j$ while discordant if $x_i < x_j$ and $y_i > y_j$

Definition 2 : Let $(X_1, Y_1$ and (X_2, Y_2) be independent random vectors, each with joint distribution function H and copulas $C_1 and C_2$. Q is defined as the difference between the probabilities of concordance and discordance of (X_1, Y_1) and (X_2, Y_2) i.e:

$$Q(C_1, C_2) = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \tag{3.2}$$

Therefore, the population version of Kendal's Tau is defined as the probability of concordance minus the probability of discordance i.e:

$$\tau_{X,Y} = P[(X_i - X_j)(Y_i - Y_j) > 0] - P[(X_i - X_j)(Y_i - Y_j) < 0] \tag{3.3}$$

Definition 3 : Let $(X_i, Y_i,$ (X_j, Y_j) and (X_k, Y_k) be independent random vectors, each with joint distribution function H whose margins are F and G. The population version of Spearman's rho is defined as the difference between probabilities of concordance and discordance of the vectors (X_i, Y_i) and (X_k, Y_k)

$$\rho_{X,Y} = 3P[(X_i - X_j)(Y_i - Y_k) > 0] - P[(X_i - X_j)(Y_i - Y_k) < 0] \tag{3.4}$$

While these two measures are an improvement to the traditional Pearson's correlation coefficient, they are still unable to fully explain the complex dependence structure between variables and capture tail dependence. Hence, this calls for a method which addresses these pertinent limitations.

3.4.2 Copulas

Copulas are joint distributions obtained from univariate marginal distributions.

Definition 4 : A two-dimensional copula function is defined as a bivariate function C$[0,1]^2 \rightarrow [0,1]$ which satisfies the following three properties:

1. (Grounded) - C(u,0) = C(0,v) = u for any u ϵ $[0,1]$
2. (With margins) - C(u,1) = C(1,u) = u for any u ϵ $[0,1]$
3. (2-increasing) - For all 0$\leq u_1 \leq u_2$ and 0$\leq v_1 \leq v_2$,

   $V([u_1, u_2] \times [v_1, v_2]) = C(u_2, v_2)$ - $C(u_1, v_2)$ - $C(u_2, v_1)$ + $C(u_1, v_{1)} \geq 0$

Sklar's Theorem: Let H be an n-dimensional distribution function with margins F_1,..... F_n, Then there exists an n-copula C such that,

$$H(x_1, ...x_n) = C(F_1(x_1), ..., F_n(x_n)) \tag{3.5}$$

The corollary to Sklar's Theorem states that: C(u,v) = H($F^{-1}(u), G^{-1}(v)$)

Sklar's theorem establishes that a joint probability distribution can be uniquely decomposed into its marginal distributions and a copula. The copula exclusively encapsulates the dependence structure between variables X and Y, disentangling it from their individual marginal behaviors. This separation is pivotal: copulas isolate how X and Y interact (e.g., tail dependencies, nonlinear associations) from their standalone statistical properties (e.g., skewness, variance). In contrast, conventional joint distribution functions conflate marginal characteristics and dependence, rendering them inseparable. By decoupling these components, copulas enable targeted analysis of interdependencies—critical for applications like risk assessment or climate modeling, where understanding how variables interact, independent of their individual scales, is paramount.

Copulas exist in various families but the common ones include: Elliptical copulas (Gaussian, Student-t) and Archimedian copulas (Clayton, Frank, Gumbel). Their functional forms are illustrated below.

For a two dimensional case, the bivariate normal (Gaussian) copula with parameter ϱ is then defined by application of Sklar's theorem:

$$C(u, v; \varrho) := \Phi_2 \left(\Phi^{-1}(u), \Phi^{-1}(v); \varrho \right) \tag{3.6}$$

where Φ^{-1} denotes the quantile function of the univariate gaussian distribution

On the other hand, the bivariate t-copula with parameters ϱ and ν is defined by:

$$C(u, v; \varrho, \nu) := T_2 \left(t_\nu^{-1}(u), t_\nu^{-1}(v); \varrho, \nu \right) \tag{3.7}$$

where t_ν^{-1} denotes the quantile function of the univariate Student's t-distribution with ν degrees of freedom.

Definition 5 : Let ψ be a continuous, strictly decreasing function from **I** to $[0, \infty]$ such that $\psi(0) = 1$, ψ^{-1} be the pseudo-inverse of ψ. We define the pseudo- inverse of ψ as the function ψ^{-1} given by:

Let $C : I^2 \rightarrow \mathbf{I}$ be given as C(u,v) = ψ^{-1}(ψ(u) + ψ(v))

Then C(u,v) is a copula known as the *Archimedian* copula. The function ψ is the generator of the copula. Some of the examples include:

- The *Gumbel* copula

It is given by the function:

$$C_\theta(u,v) = \exp[-(-logu)^\theta + (-logv)^\theta)^{\frac{1}{\theta}}] \qquad (3.8)$$

whereby $\theta \epsilon$ $[1, \infty)$.

The generator for this copula is the function

$$\psi \text{ (t)} = (-logt)^\theta$$

- The *Clayton* copula

It is given by the function

$$C_\theta(u,v) = [max(u^{-\theta} + v^{-\theta} - 1, 0)]^{\frac{-1}{\theta}} \qquad (3.9)$$

for $\theta \epsilon$ [-1, ∞)|0.

The generator for this copula is the function:

$$\psi(t) = \frac{1}{\theta}(t^{-\theta} - 1)$$

- The *Frank* copula

It is given by the function

$$C_\theta(u,v) = -\frac{1}{\theta} ln(1 + \frac{(\exp^{-\theta u} - 1)(\exp^{-\theta v} - 1)}{\exp^{-\theta} - 1}) \tag{3.10}$$

for θ ϵ (-∞, ∞) | 0).

The generator for this copula is the function

$$-ln \frac{\exp^{-\theta t} - 1}{\exp^{-\theta} - 1}$$

It is possible to derive expressions for the conditional density and distribution functions, which are important later. In particular, the conditional density $f_{1|2}$ and distribution function $F_{1|2}$ can be expressed as:

$$f_{1|2}(x_1|x_2) = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_2(x_2) \tag{3.11}$$

$$F_{1|2}(x_1|x_2) = \frac{\partial}{\partial F_2(x_2)} C_{12}(F_1(x_1), F_2(x_2)) = \frac{\partial}{\partial v} C_{12}(F_1(x_1), v)\Big|_{v=F_2(x_2)} \tag{3.12}$$

A trivariate copula is a statistical model that describes the joint dependence structure between three variables, separating their dependence structure from their individual marginal distributions. It is used to model and study the dependence structure between three variables while keeping their individual (marginal) distributions separate. Trivariate copula models capture how three variables relate to each other in terms of their dependence. This includes not just linear relationships (correlation) but also more complex dependencies, such as tail dependence (extreme values occurring together).

For instance, in the context of maize production, rainfall, and temperature, a trivariate copula can help model how extreme weather events (like droughts or heatwaves) might jointly impact crop yield. A trivariate copula focuses on the joint dependence, separate from the marginal distributions of the variables. This allows one to model each variable with its own distribution (e.g., normal, skewed) while studying their joint behaviour. The marginal distributions are often transformed to uniform distributions on the interval [0, 1] using their empirical distribution functions before applying a copula.

The tool of copulas is less universal in the case of m ($m \geq 3$) variables than it is in the case of two. However, an m-dimensional copula function C_H can be constructed from an m-dimensional cdf H with margins $H_1, \ldots, H_m$ as

$$C_H(u_1, \ldots, u_m) = H\left(H_1^{-1}(u_1), \ldots, H_m^{-1}(u_m)\right), \quad (u_1, \ldots, u_m) \in [0,1]^m. \tag{3.13}$$

In the case $m = 3$, choosing as H the three-dimensional Gaussian cdf Φ_R with standard normal marginals and covariance matrix equal to the correlation matrix

$$R = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix},$$

we get the following trivariate Gaussian copula:

$$C_R(u_1, u_2, u_3) = \Phi_R(q_1, q_2, q_3) = \int_{-\infty}^{q_1} \int_{-\infty}^{q_2} \int_{-\infty}^{q_3} \phi_R(x, y, z)\, dx\, dy\, dz \tag{3.14}$$

where

$$\phi_R(x, y, z) = \frac{1}{(2\pi)^{3/2}|R|^{1/2}} \exp\left(-\frac{1}{2}(x, y, z) R^{-1} (x, y, z)^\top\right)$$

is the three-dimensional Gaussian pdf with zero mean and covariance matrix R, and $q_k = \Phi^{-1}(u_k)$ for $k \in \{1, 2, 3\}$ are the normal scores.

The density of this copula is:

$$c_R(u_1, u_2, u_3) = \frac{\partial^3 C_R(u_1, u_2, u_3)}{\partial u_1 \partial u_2 \partial u_3} = \frac{\phi_R(q_1, q_2, q_3)}{\phi(q_1)\phi(q_2)\phi(q_3)} = \frac{1}{|R|^{1/2}} \exp\left(\frac{1}{2}\mathbf{q}^\top (I - R)\mathbf{q}\right)$$

where $\mathbf{q} = (q_1, q_2, q_3)^\top$ is the vector of normal scores, I is the 3×3 identity matrix, and ϕ is the standard univariate normal pdf.

To define the multivariate t copula, suppose there are m variates, and $\mathbf{u}$ is a vector of m probability values (numbers in $[0, 1]$). Let $\mathbf{s}$ be the vector of the univariate t-quantiles of $\mathbf{u}$ with n degrees of freedom, that is $s_i = F_n^{-1}(u_i)$ for each element of $\mathbf{s}$ and $\mathbf{u}$. Also let Σ be an $m \times m$ correlation matrix with determinant d. Where m is 3, the t-copula has density:

$$c(\mathbf{u}; n, \Sigma) = K_3 \frac{\left[\prod_{i=1}^{3}\left(1 + \frac{s_i^2}{n}\right)\right]^{(n+1)/2}}{\left(1 + \frac{\mathbf{s}^\top \Sigma^{-1} \mathbf{s}}{n}\right)^{(3+n)/2}}, \tag{3.15}$$

where

$$K_m = \frac{\Gamma\left(\frac{3+n}{2}\right)[\Gamma(n/2)]^2}{[\Gamma(\frac{1}{2} + \frac{n}{2})]^3 d^{1/2}}.$$

By starting with a Kendall's τ coefficient matrix $\mathbf{T}$, the correlation matrix needed here can be specified by

$$\Sigma = \sin\left(\frac{\mathbf{T}\pi}{2}\right).$$

For Archimedean copulas like Gumbel, Clayton and Frank, the trivariate copula can be obtained by extending to an order greater than 2 using the property of Archimedean copulas.

A trivariate Gumbel copula is expressed as:

$$C_\theta(u1, u2, u3) = \exp[-(-logu1)^\theta + (-logu2)^\theta + (-logu3)^\theta)^{\frac{1}{\theta}}] \tag{3.16}$$

whereby $\theta \, \epsilon$ [1, ∞).

A trivariate Clayton copula is given as:

$$C_\theta(u1, u2, u3) = [max(u1^{-1/\theta} + u2^{-1/\theta} + u3^{-1/\theta} - 2, 0)]^{-\theta} \tag{3.17}$$

for $\theta \, \epsilon$ [-1, ∞)|0.

A trivariate Frank Copula is given by:

$$C_\theta(u1, u2, u3) = -\frac{1}{\theta} ln(1 + \frac{(\exp^{-\theta u1} - 1)(\exp^{-\theta u2} - 1)(\exp^{-\theta u3} - 1)}{\exp^{-\theta} - 1}) \tag{3.18}$$

for θ ϵ (-∞, ∞) | 0).

In addition to correlation coefficients, bivariate and trivariate copula models, this study incorporated vine copula models. Following Czado and Nagler (2022), the motivation for vine copula models was to find a way to construct multivariate copulas using only bivariate copulas as building blocks. The appropriate tool to obtain such a construction is conditioning. Joe (1996) gave the first pair copula construction in terms of distribution functions, while Bedford & Cooke (2001) and Bedford & Cooke (2002) independently developed constructions in terms of densities. They also provided a framework to identify all possible constructions. We first illustrate this construction for $d = 3$ by starting with the recursive factorization:

$$f(x_1, x_2, x_3) = f_{3|12}(x_3|x_1, x_2) f_{2|1}(x_2|x_1) f_1(x_1) \tag{3.19}$$

and treat each part separately. Here $F_{j|D}$ and $f_{j|D}$ denote the conditional distribution or density function of X_j given X_D, respectively. To determine $f_{3|12}(x_3|x_1, x_2)$, we consider the bivariate conditional density $f_{13|2}(x_1, x_3|x_2)$. The copula density $c_{13;2}(\cdot, \cdot; x_2)$ denotes the copula density associated with the conditional distribution of (X_1, X_3) given $X_2 = x_2$. Using eqn 3.5 for $f_{13|2}(x_1, x_3|x_2)$ gives:

$$f_{13|2}(x_1, x_3|x_2) = c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) f_{1|2}(x_1|x_2) f_{3|2}(x_3|x_2) \tag{3.20}$$

Now $f_{3|12}(x_3|x_1, x_2)$ is the conditional density of X_3 given $X_1 = x_1$, $X_2 = x_2$, which can be determined using Equation 3.11 applied to Equation 3.20, yielding:

$$f_{3|12}(x_3|x_1, x_2) = c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2) f_{3|2}(x_3|x_2). \tag{3.21}$$

Finally, direct application of Equation 3.12 gives:

$$f_{2|1}(x_2|x_1) = c_{12}(F_1(x_1), F_2(x_2))f_2(x_2)f_{3|2}(x_3|x_2) \quad = c_{23}(F_2(x_2), F_3(x_3))f_3(x_3) \tag{3.22}$$

Inserting Equations 3.21-3.23 into Equation 3.19 yields a pair copula decomposition of an arbitrary three-dimensional density $f(x_1, x_2, x_3)$ as:

$$f(x_1, x_2, x_3) = c_{13;2}\big(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); x_2\big) \cdot c_{23}\big(F_2(x_2), F_3(x_3)\big) \cdot c_{12}\big(F_1(x_1), F_2(x_2)\big) f_3(x_3)f_2(x_2)f_1(x_1). \tag{10}$$

We see that the joint density can be expressed in terms of bivariate copula densities, marginal densities, and conditional distribution functions. However, this decomposition is not unique:

$$f(x_1, x_2, x_3) = c_{23;1}\big(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1); x_1\big) \cdot c_{13}\big(F_1(x_1), F_3(x_3)\big) \cdot c_{12}\big(F_1(x_1), F_2(x_2)\big) f_3(x_3)f_2(x_2)f_1(x_1) \tag{3.23}$$

$$f(x_1, x_2, x_3) = c_{12;3}\big(F_{1|3}(x_1|x_3), F_{2|1}(x_2|x_1); x_3\big) \cdot c_{13}\big(F_1(x_1), F_3(x_3)\big) \cdot c_{23}\big(F_2(x_2), F_3(x_3)\big) f_3(x_3)f_2(x_2)f_1(x_1) \tag{3.24}$$

are two different decompositions using a reordering of the variables in Equation 3.19.

All decompositions of the density have a conditional copula term of the form $c_{ij;k}(\cdot, \cdot; x_k)$, called pair copula. To facilitate estimation we normally neglect the dependence on the specific conditioning value x_k. This is called the *simplifying assumption.* For a three-dimensional density with copula parameter vector $\boldsymbol{\theta} = (\theta_{12}, \theta_{23}, \theta_{13;2})$, we then get the following simplified pair copula construction:

$$f(x_1, x_2, x_3; \boldsymbol{\theta}) = c_{13;2}\big(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); \theta_{13;2}\big) \cdot c_{23}\big(F_2(x_2), F_3(x_3); \theta_{23}\big) \cdot c_{12}\big(F_1(x_1), F_2(x_2); \theta_{12}\big) f_3(x_3)f_2(x_2)f_1(x_1) \tag{3.25}$$

where $c_{13;2}(\cdot, \cdot; \theta_{13;2})$, $c_{12}(\cdot, \cdot; \theta_{12})$, and $c_{23}(\cdot, \cdot; \theta_{23})$ are arbitrary parametric bivariate copula densities. The dependence on marginal parameters has been suppressed to ease notation. This is no longer a decomposition but a construction, where the dependence on x_2 in $c_{13;2}(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2); \theta_{13;2})$ is solely captured by the arguments. If the margins in Equation 13 are uniform, we have a three-dimensional parametric copula density.

For the estimation of the parameter $\boldsymbol{\theta}$ based on an i.i.d. sample $\mathbf{x}_k = (x_{k1}, x_{k2}, x_{k3})$, $k = 1, \dots, n$, we follow the two-step approach discussed in Section 1. We create the associated pseudo-data $u_{k,j}$, $k = 1, \dots, n$, $j = 1, \dots, 3$. This allows us to write the joint (pseudo-)likelihood for the trivariate copula density associated with Equation 3.25 as:

$$\ell(\boldsymbol{\theta}; \mathbf{u}) = \prod_{k=1}^{n} c_{13;2}(C_{1|2}(u_{k,1}|u_{k,2}; \theta_{12}), C_{3|2}(u_{k,3}|u_{k,2}; \theta_{23}); \theta_{13;2}) \cdot c_{23}(u_{k,2}, u_{k,3}; \theta_{23}) c_{12}(u_{k,1}, u_{k,2}; \theta_{12}). \tag{3.26}$$

Maximizing Equation 3.26 gives the joint maximum likelihood estimator $\hat{\boldsymbol{\theta}}$. However, there is an alternative sequential estimation method, which remains computationally tractable in high dimensions. First, we find parameter estimates $\hat{\theta}_{12}$ and $\hat{\theta}_{23}$ by maximizing:

$$\prod_{k=1}^{n} c_{12}(u_{k,1}, u_{k,2}; \theta_{12}) \quad \text{and} \quad \prod_{k=1}^{n} c_{23}(u_{k,2}, u_{k,3}; \theta_{23})$$

over θ_{12} and θ_{23}, respectively. Second, we define the pseudo-observations:

$$u_{k,1|2;\hat{\theta}_{12}} = C_{1|2}(u_{k,1}|u_{k,2}; \hat{\theta}_{12}) \quad \text{and} \quad u_{k,3|2;\hat{\theta}_{23}} = C_{3|2}(u_{k,3}|u_{k,2}; \hat{\theta}_{23}) \tag{3.27}$$

for $k = 1, \dots, n$. Under the simplifying assumption, these provide an approximate i.i.d. sample from the pair copula $C_{13;2}$. Further, the marginal distribution associated with the pseudo-observations Equation 3.27 is approximately uniform, since the transformation in Equation 3.27 is a probability integral transform with estimated parameter values. Therefore, we use them to estimate the parameter(s) of the pair copula $c_{13;2}$ by maximizing:

$$\prod_{k=1}^{n} c_{13;2}(u_{k,1|2;\hat{\theta}_{12}}, u_{k,3|2;\hat{\theta}_{23}}; \theta_{13;2})$$

over $\theta_{13;2}$. This splits the estimation of $\boldsymbol{\theta}$ into three simpler problems. The sequential estimate can also be used as a starting value for the joint maximum likelihood estimation. A similar sequential approach can also be followed when estimating pair copulas nonparametrically.

3.4.3 Semiparametric Estimation

Semiparametric estimation combines parametric and nonparametric approaches. Li et al. (2019a) describes a semi-parametric approach as when the marginal distribution is estimated non-parametrically without making assumptions about the data's parametric form, but the copula model is parametrically estimated using methods such as maximum likelihood. The semi-parametric approach, also known as the Canonical Maximum Likelihood method involves first transforming the marginal data by the empirical distribution function into uniformed variates without making any assumptions on the parametric form for each of them. Next, holding these marginals fixed, the copula dependence parameter is estimated by maximizing the log-likelihood function.

While parametric methods excel in simplicity, their rigidity undermines reliability in complex, non-Gaussian systems like agriculture, semiparametric estimation balances flexibility and structure, making it ideal for agro-climatic studies. This framework avoids biases from assuming incorrect parametric forms Chen et al. (2006); retains interpretability in modeling dependencies (e.g., tail risks), accommodates shifting climatic baselines (e.g., rising temperatures) and evolving dependence patterns Tschora and Cherubini (2020).

3.4.4 Data analysis

Data analysis was performed using the R open source software.

Chapter 4: Results and Discussion

4.1 Introduction

This chapter presents the descriptive statistics of the collected dataset, along with the analysis results and a comprehensive discussion of these findings. The section commences with an overview of the summary statistics of the data, followed by an in-depth exploration of copula fitting to model the dependence structures within the dataset.

4.2 Descriptive statistics

This section presents the summary statistics and time series plots for maize production, temperature and precipitation.

Table 4.1: Summary statistics for maize production, precipitation and temperature

Variable	Min	Max	Median	Mean	Std. dev
Maize production	101.82	425.14	254.51	249.45	102.35
Temperature	17.32	18.21	17.71	17.74	0.2
Precipitation	83.33	216.33	164.58	161.23	27.26

Table 4.2: Summary statistics for area under production and yield per ha

Variable	Min	Max	Median	Mean	Std. dev
Area under production	41.8	138.14	65.76	74.15	4.09
Yield per ha	1.94	4.96	3.15	3.29	0.12

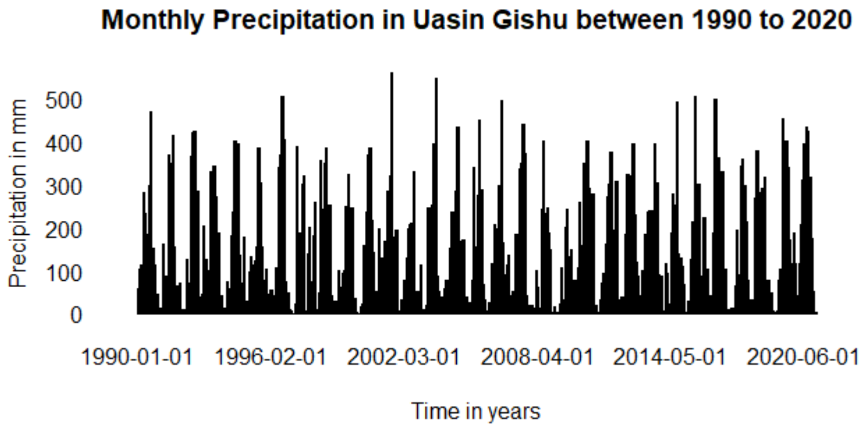


Figure 4.1: Monthly Precipitation in Uasin Gishu between 1990 to 2020

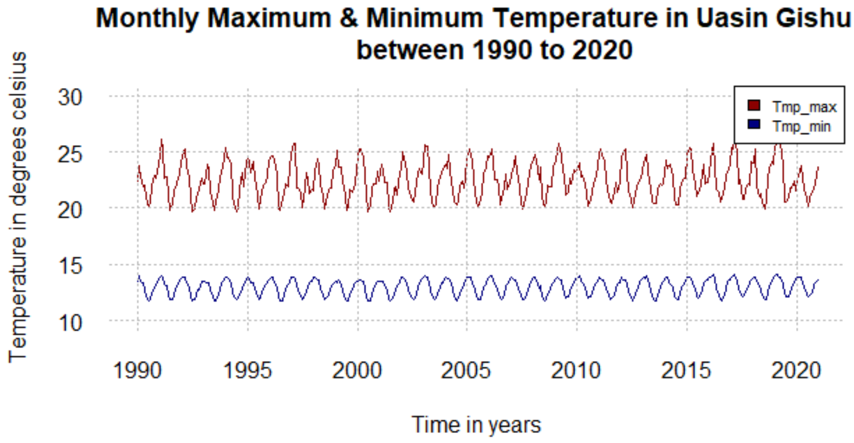


Figure 4.2: Monthly Maximum and Minimum Temperature in Uasin Gishu between 1990 to 2020

The first step was to get a general understanding of the data using basic statistics such as mean, median and standard deviation of the variables under study. Area (Ha per 000) represents the area of maize cultivation in hectares per 1000 units. It was noted that there was a significant increase in the hectares of land used for maize production. Up from 61.238 ha in the year 1990 to 138.14 ha in 2020, this represents a remarkable surge of 126%. This could be attributed to several factors such as government incentives to farmers through seed and fertilizer subsidies, favourable market prices which attracted the farmers to increase their acreage under cultivation or simply a growing population within the region thus necessitating elevated demand for consumption and increased production.

Thanks to advanced mechanization, improved quality of inputs and other factors, maize production grew from an average of about 160.34 tonnes in 90s to approximate average of 350.69 tonnes in the between 2010 and 2020. Similarly, the Yield (Tons/Ha) which indicates the yield of maize per hectare improved from 1.939 tons per hectare in 1990 to 4.960 tons per hectare in 2020, a figure that stands way above the national mean yield of 2 metric tons per hectare. It is also worth noting that there's an apparent positive correlation between the rise in area under maize production and the production and yield over the years under study. Equally, the minimum production was recorded in the year 1997 which coincides to be the time when the country experienced the El Nino rains. Above average rains can cause water logging, sweeping away planted seeds and destroying young plants thus reduced yields which could possibly inform why maize production was very low.

A look at the annual average temperature reveals that in the years 2009, 2015, 2017 and 2019, the temperatures were above 18 degrees Celsius. An occurrence of about four times in the 2009 to 2019 decade possibly points to an emerging trend of increasing temperatures in the region. This could be due to the urbanization effect as industries are set up in the county leading to high levels of carbon emissions in the recent past. Although the changes above the mean of 17.7 degrees Celsius is seemingly negligible, if it is above the threshold required for the normal growth of crops, it will affect the germination, flowering and maturity of agricultural produce.

At the same time, precipitation has seen an upwards trend between 1990 and 2020 apart from a few years such as 1999, 2002, 2008, 2009 and 2017 whereby the precipitation was way below the mean of about 161 mm. Quite interestingly, the year 2009 experienced the lowest rainfall amounting to just 83 mm. Coincidentally, the same year recorded the highest temperature during the study period. This is clearly attributed to the drought which affected countries in East Africa between 2008 and early 2010.

The data only revealed three months when no rainfall was recorded i.e. February 1997, February 2000 and January 2012. These data points accounted for just less than 1% of the total observations. Monthly precipitation within the 30 year period was fairly within the mean of 161.2mm. However, there were instances when the monthly recorded amounts went as high as 561mm in August 2001. As expected, the months of June, July and August experienced a lot of rainfall for most of the years. Temperatures averaged lows of 12 degrees Celsius to highs of 26 degrees Celsius with a mean of about 17.7 degrees Celsius over the 30-year period.

4.3 Correlation analysis

This section presents correlation analysis results for maize production, precipitation and temperature using Perason's correlation coefficient, Spearman's rho and Kendall's tau.

Table 4.3: Correlation matrix for maize production, precipitation, temperature using Pearson's coefficient

Variables	Maize production	Temperature	Precipitation
Maize production	1	0.24	0.26
Temperature	0.24	1	−0.52
Precipitation	0.26	−0.52	1

Table 4.4: Correlation matrix for maize production, precipitation, temperature using Spearman's rho

Variables	Maize production	Temperature	Precipitation
Maize production	1	0.08	0.29
Temperature	0.08	1	−0.52
Precipitation	0.29	−0.52	1

Table 4.5: Correlation matrix for maize production, precipitation, temperature using Kendall's tau

Variables	Maize production	Temperature	Precipitation
Maize production	1	0.05	0.18
Temperature	0.05	1	-0.38
Precipitation	0.18	-0.38	1

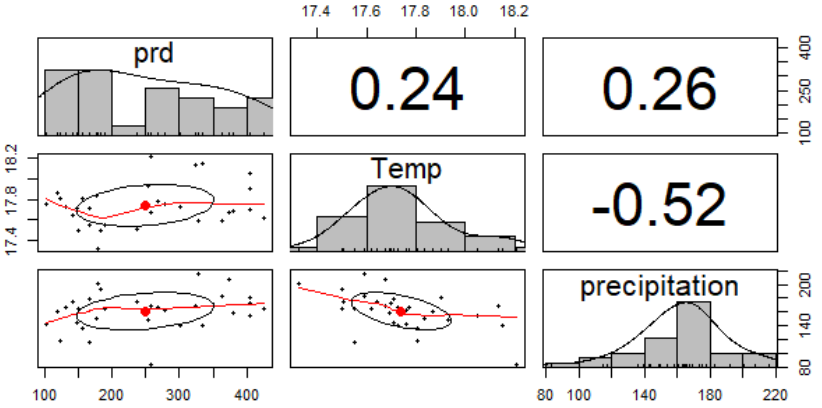


Figure 4.3: Pair plot showing correlation plots, histograms and correlation matrix for maize production, precipitation, temperature using Pearson's correlation coefficient

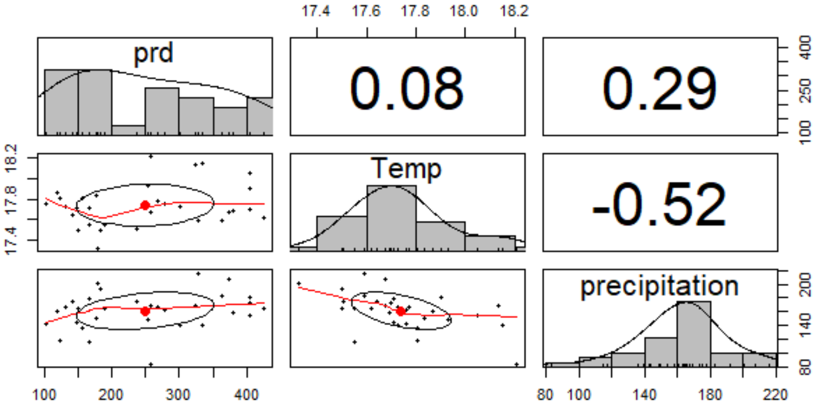


Figure 4.4: Pair plot showing correlation plots, histograms and correlation matrix for maize production, precipitation, temperature using Spearman's rho

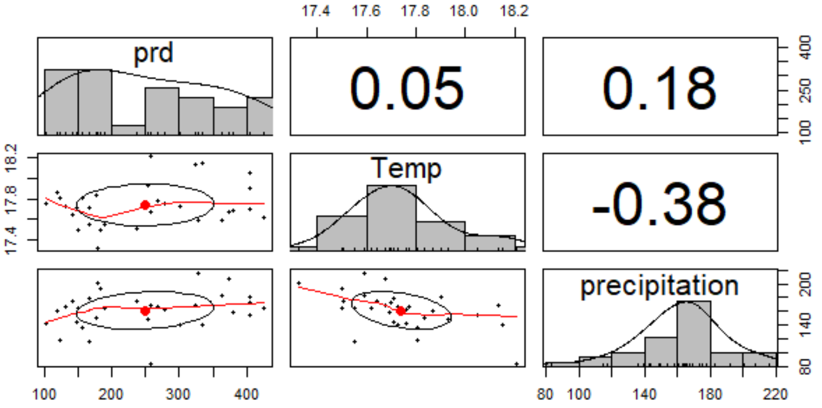


Figure 4.5: Pair plot showing correlation plots, histograms and correlation matrix for maize production, precipitation, temperature using Kendall's tau

Table 4.6: Correlation matrix for detrended maize production, precipitation, temperature using Pearson's coefficient

Variables	Maize production	Temperature	Precipitation
Maize production	1	-0.15	0.25
Temperature	-0.15	1	-0.52
Precipitation	0.25	-0.52	1

Table 4.7: Correlation matrix for detrended maize production, precipitation, temperature using Spearman's rho

Variables	Maize production	Temperature	Precipitation
Maize production	1	-0.14	0.26
Temperature	-0.14	1	-0.52
Precipitation	0.26	-0.52	1

Table 4.8: Correlation matrix for detrended maize production, precipitation, temperature using Kendall's tau

Variables	Maize production	Temperature	Precipitation
Maize production	1	-0.12	0.19
Temperature	-0.12	1	-0.38
Precipitation	0.19	-0.38	1

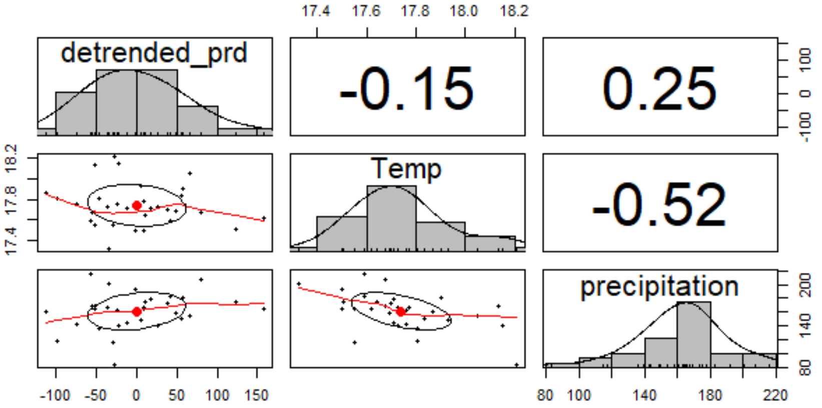


Figure 4.6: Pair plot showing correlation plots, histograms and correlation matrix for maize production, precipitation, temperature using Pearson's correlation coefficient

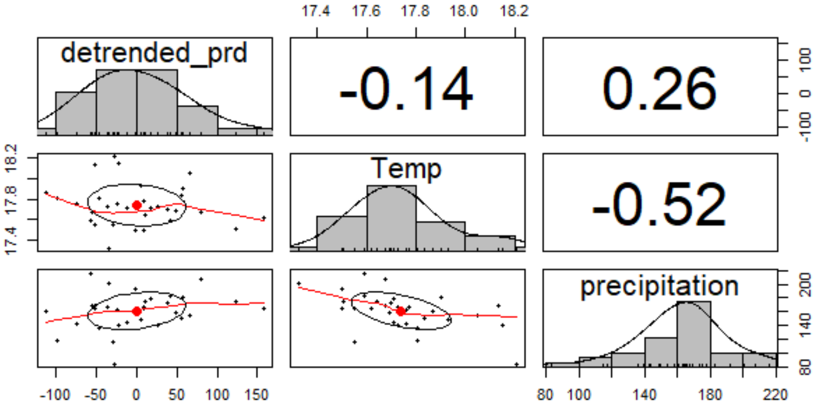


Figure 4.7: Pair plot showing correlation plots, histograms and correlation matrix for maize production, precipitation, temperature using Spearman's rho

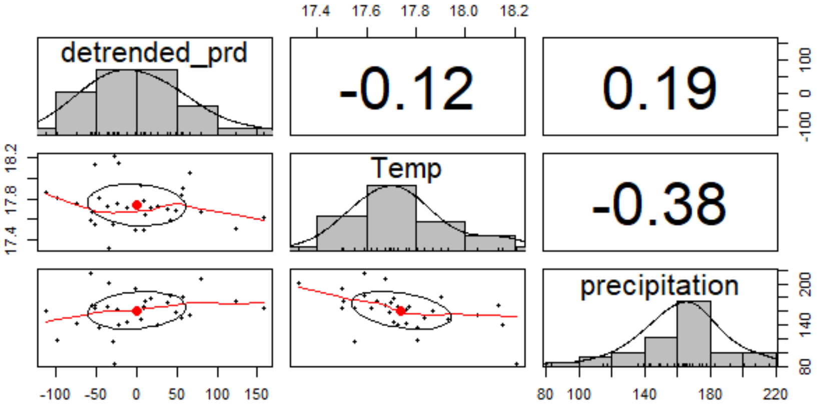


Figure 4.8: Pair plot showing correlation plots, histograms and correlation matrix for maize production, precipitation, temperature using Kendall's tau

The Pearson correlation coefficient between precipitation and temperature was found to be equal to- 0.52. This indicates a moderate negative correlation, suggesting that higher temperatures are associated with lower precipitation levels. The Spearman's rho coefficient is also -0.52, indicating a moderate negative relationship but based on rank rather than the actual values, which accounts for non-linear relationships. The Kendall's tau coefficient is - 0.38, showing a moderate negative association like Spearman's but with a measure of concordance and discordance in pairs of observations. The negative correlation between precipitation and temperature suggests that climatic changes affecting temperature are likely to have significant impacts on precipitation patterns and vice versa.

The Pearson correlation coefficient between maize production and temperature and maize production was equal to 0.24. This indicates a weak positive correlation, suggesting that there is a slight tendency for higher temperatures to be associated with increased maize production. The Spearman's rho coefficient is 0.08, indicating a very weak positive rank-based relationship, suggesting negligible correlation. The Kendall's tau coefficient is 0.05, also indicating a negligible positive association. Spearman's and Kendall's measures suggest an even weaker relationship than Pearson's, indicating that temperature variations have minimal impact on maize production. However, Pearson correlation coefficient between detrended maize production and temperature was - 0.15, Spearman's rho coefficient - 0.14 and Kendall's tau coefficient - 0.12. Detrending the maize production dataset uncovered a weak negative relationship compared to the weak positive relationship when the original maize production dataset is used. The weak negative correlations between detrended maize production and temperature suggest that temperature is a limiting factor to maize production. This trend is consistent with reviewed literature.

The Pearson correlation coefficient between maize production and precipitation was 0.26. This indicates a weak positive correlation, suggesting that higher precipitation levels are associated with a slight increase in maize production. The Spearman's rho coefficient is 0.29, indicating a slightly stronger positive relationship when considering rank. The Kendall's tau coefficient is 0.18, indicating a weak positive association, but less strong compared to Pearson's and Spearman's results. On the other hand, Pearson correlation coefficient between detrended maize production and precipitation was 0.25, Spearman's rho coefficient 0.26 and Kendall's tau coefficient 0.19. This indicates a weak positive correlation, suggesting that higher precipitation levels are associated with a slight increase in maize production. The Spearman's rho coefficient is highest, indicating a slightly stronger positive relationship when considering ranks. The relationship is weakly positive across all measures, with Spearman's rho showing slightly stronger correlations than Pearson's coefficient and Kendall's tau. These results may imply that while precipitation is an important variable in maize production, it is not the sole determinant. Other variables such as soil fertility, agricultural practices, temperature, and technological advancements in farming likely play more significant roles in influencing maize production. The weak positive correlation indicates that maize production may benefit slightly from increased precipitation, but the overall impact is limited when considering the multifaceted nature of agricultural productivity.

In general, these findings underscore the complex interplay between climatic variables and agricultural outputs, highlighting the need for multifaceted approaches to understanding and optimizing crop production under varying climatic conditions. The findings also provide a comprehensive understanding of the relationships between climatic variables and maize production, highlighting the importance of using multiple correlation measures to capture the full nature of these relationships.

4.4 Autocorrelation

This section presents the autocorrelation test results for maize production, precipitation and temperature using Ljung's box, autocorrelation (acf) and partial autocorrelation (pacf) plots.

Table 4.9: Test for autocorrelation for detrended maize production, precipitation, temperature using Box-Ljung test

Variables	Test statistic (X-squared)	Degrees of freedom	p-value
Maize production	21.739	1	0.00000312
Squared maize production	18.056	1	0.0000214
Detrended maize production	4.7466	1	0.02936
Squared detrended maize production	0.066326	1	0.7968
Temperature	0.1749	1	0.6758
Squared temperature	0.17329	1	0.6772
Precipitation	2.7257	1	0.09875
Squared precipitation	3.1518	1	0.07584

Testing for autocorrelation and stationarity was important before further analysis using copula models because autocorrelation can introduce spurious dependence structures in the data. If the autocorrelation is not properly accounted for, the copula model might capture the temporal dependence rather than the true dependence structure between variables. Temporal dependence can lead to misleading conclusions about the strength and nature of the dependence between variables. Given that copula models assume that the data are independent and identically distributed, autocorrelation violates this assumption, resulting in biased parameter estimates.

Ljung's box test is used as confirmatory test for the presence of significant autocorrelation. It assumes a null hypothesis of no autocorrelation in a dataset and an alternative hypothesis of the presence of autocorrelation in a dataset. Before detrending, the p-value was 0.000003 which indicated strong evidence of autocorrelation in the raw maize production data. However, the detrended maize production series yields a p-value of 0.02936 which is a significant improvement. Although, the p-value is not statistically significant at 1%, it is significant at 5%. It is assumed that the possible residual autocorrelation will have a negligible effect on copula modeling hence no further work done. The Ljung-box test p-values for both original and squared temperature/precipitation data were not statistically significant at 5% to indicate presence of autocorrelation.

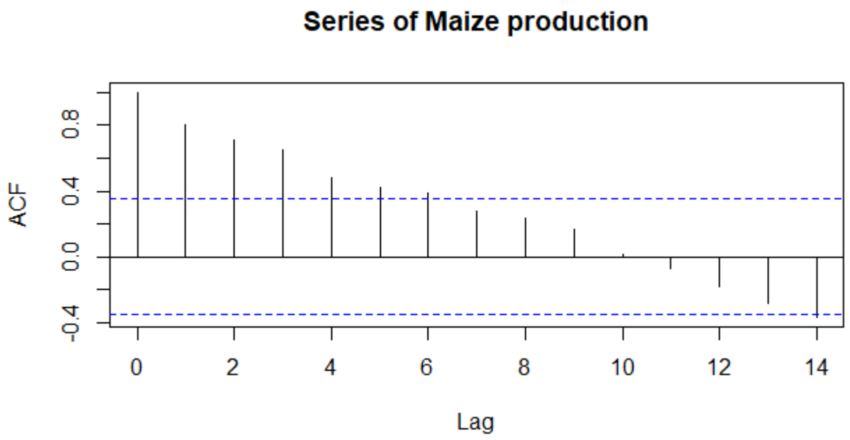


Figure 4.9: ACF plot for series of annual maize production

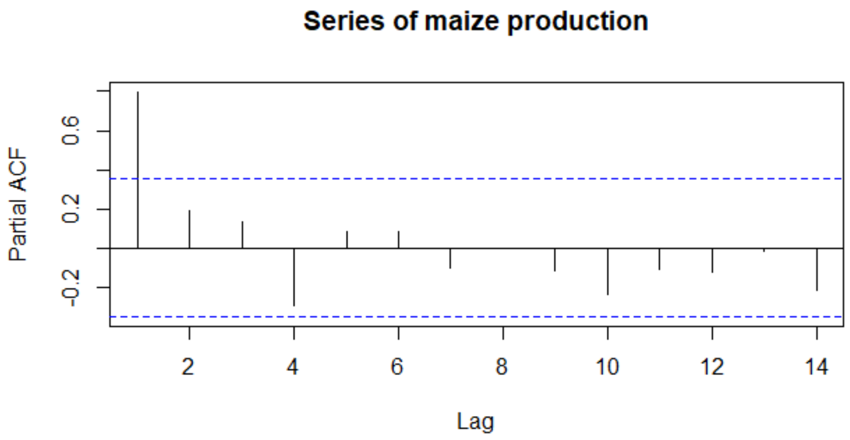


Figure 4.10: PACF plot for series of annual maize production

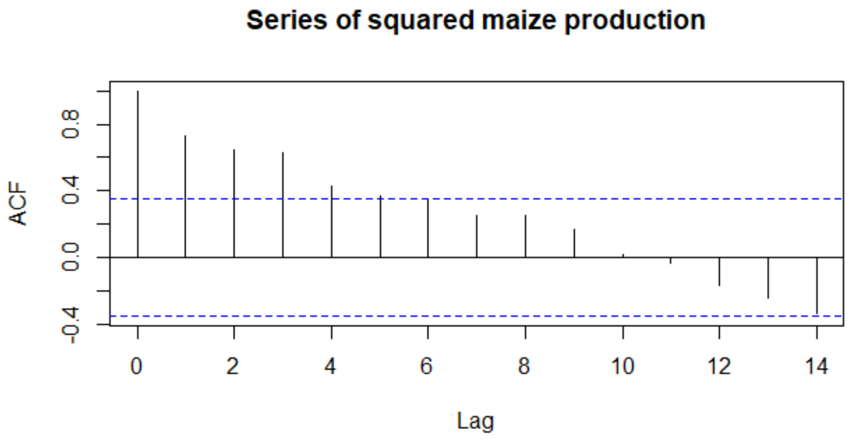


Figure 4.11: ACF plot for series of squared maize production

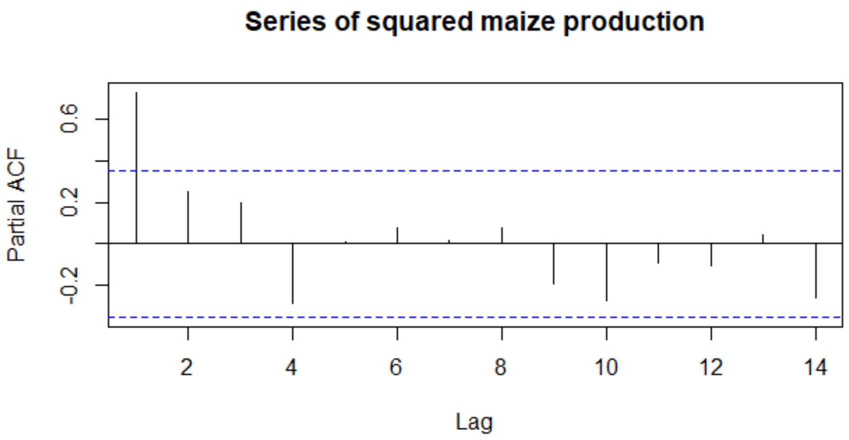


Figure 4.12: PACF plot for series of squared maize production

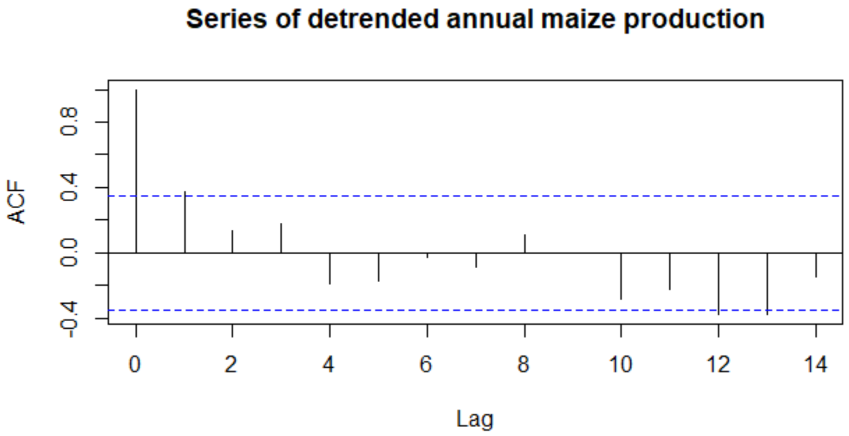


Figure 4.13: ACF plot for series of detrended annual maize production

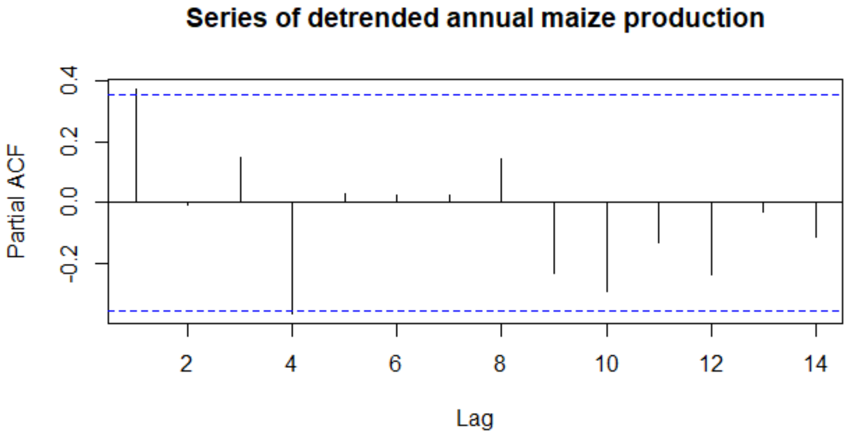


Figure 4.14: PACF plot for series of detrended annual maize production

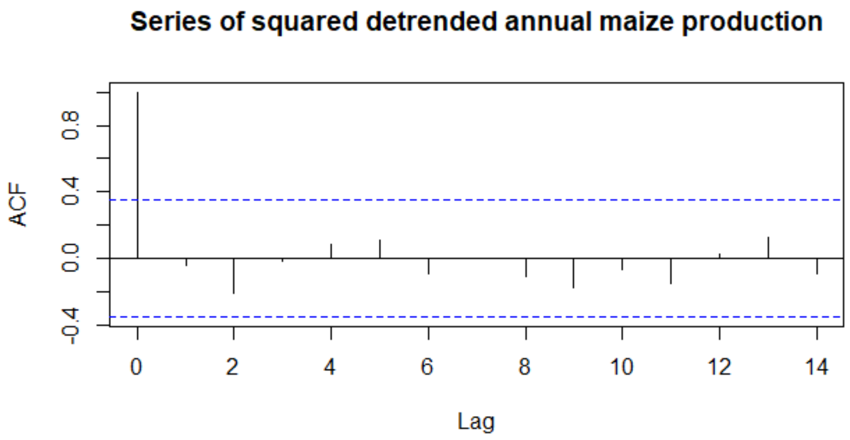


Figure 4.15: ACF plot for series of squared detrended annual maize production

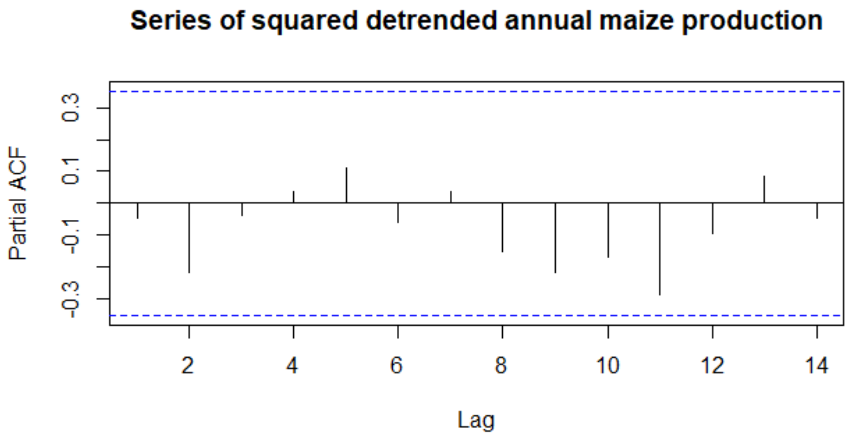


Figure 4.16: PACF plot for series of squared detrended annual maize production

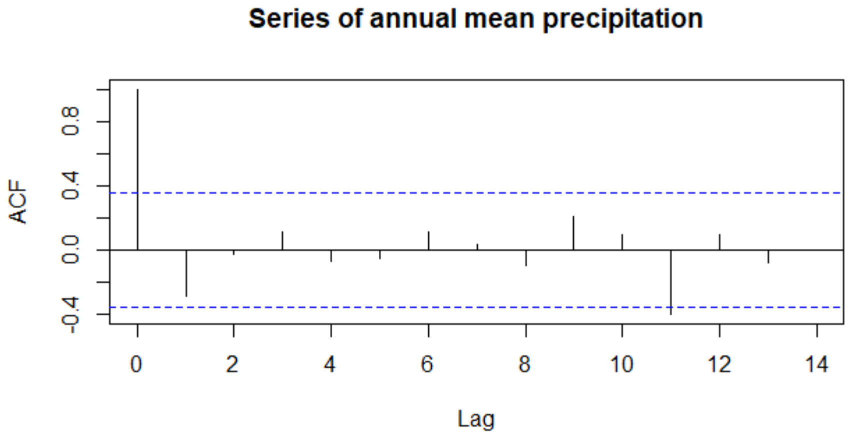


Figure 4.17: ACF plot for series of annual mean precipitation

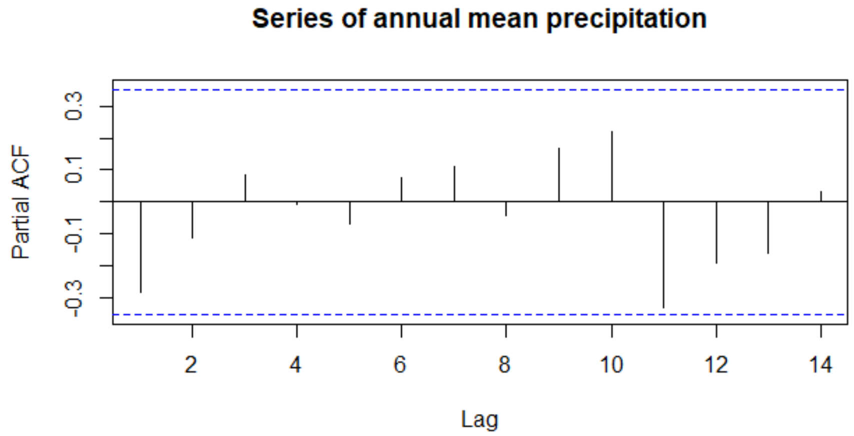


Figure 4.18: PACF plot for series of annual mean precipitation

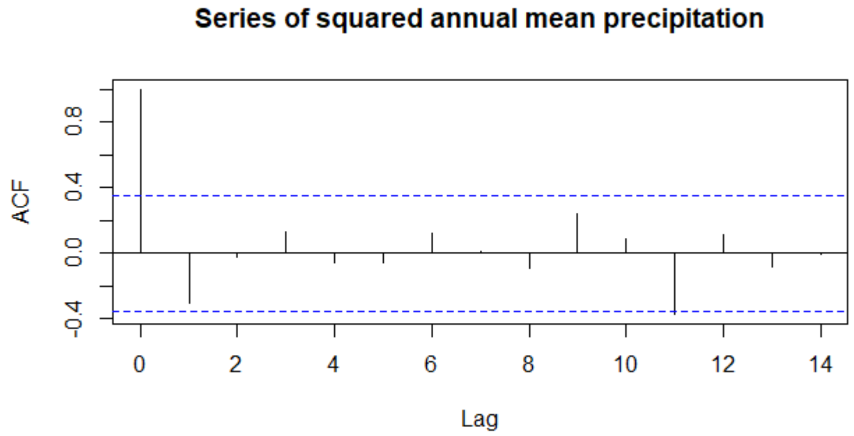


Figure 4.19: ACF plot for series of squared annual mean precipitation

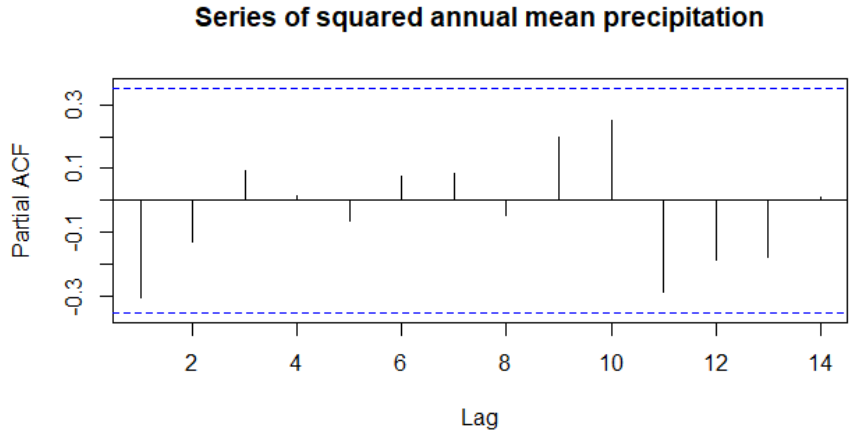


Figure 4.20: PACF plot for series of squared annual mean precipitation

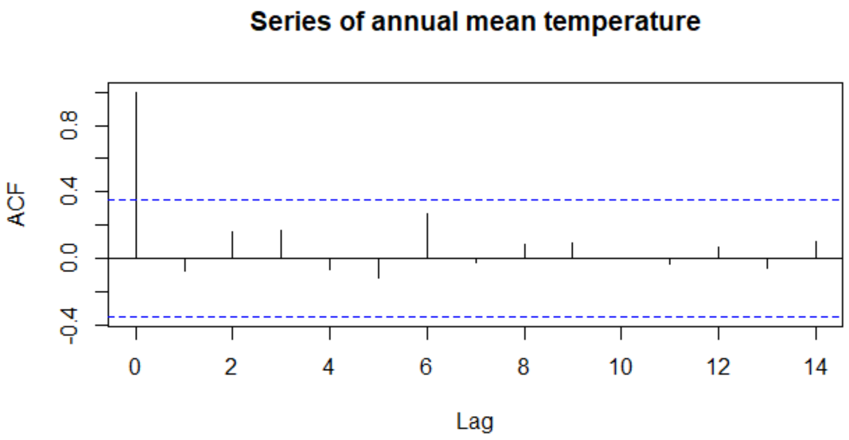


Figure 4.21: ACF plot for series of annual mean temperature

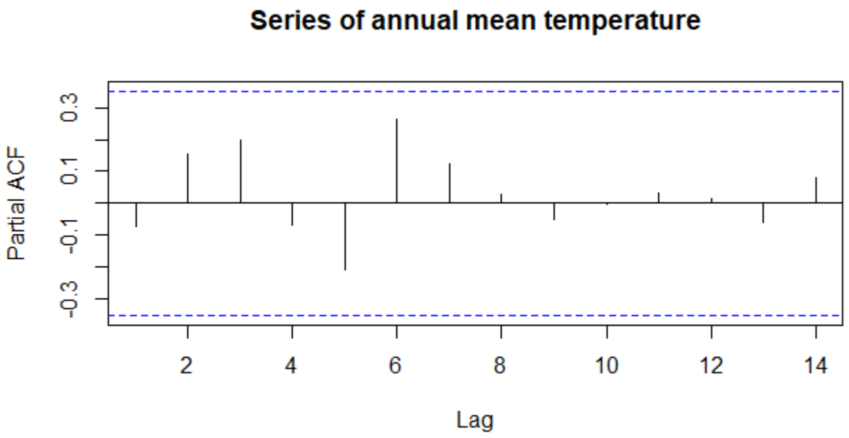


Figure 4.22: PACF plot for series of annual mean temperature

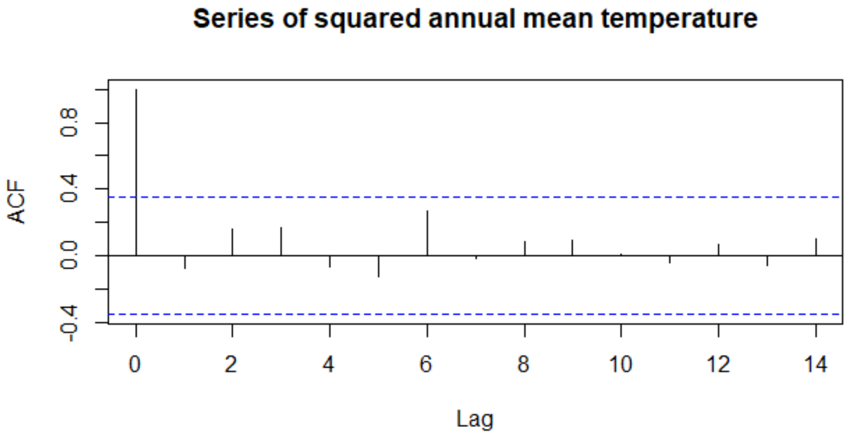


Figure 4.23: ACF plot for series of squared annual mean temperature

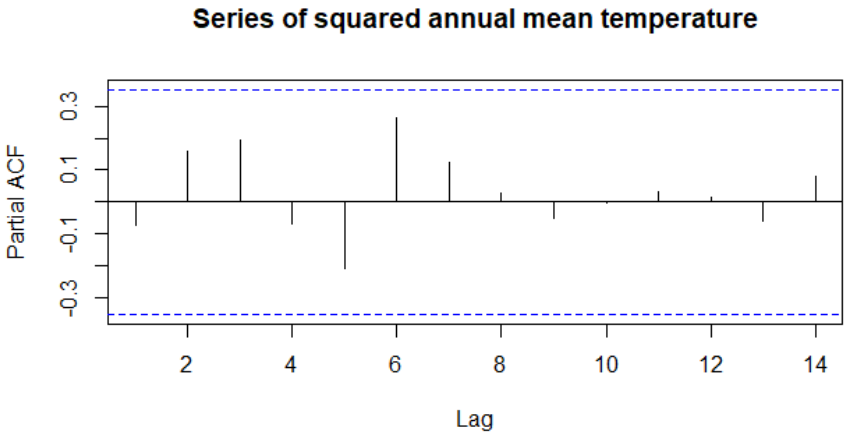


Figure 4.24: PACF plot for series of squared annual mean temperature

The variables maize production, temperature and precipitation were analysed using the R functions acf and pacf which are both tools used in time series analysis to understand the relationships between observations in the series at different lags. It was noted that the ACF plot for maize production time series exhibits significant autocorrelation at various lags, indicating a strong persistence in the data. Between Lag 1 to Lag 6, the autocorrelation coefficients were significantly positive, indicating a high degree of correlation between the current value and its recent past values. This suggested a strong trend or persistence in maize production over short time periods. The autocorrelation decreased gradually between Lag 7 to Lag 10, rather than dropping off sharply, suggesting that the series has a long memory and that past values continue to influence future values over several time periods. Between Lag 11 to Lag 14, there was slightly significant negative autocorrelations, indicating potential cyclic behaviour or seasonality in the maize production data.

Similarly, the partial autocorrelation at lag 1 was significantly positive (approximately 0.8), indicating a strong direct relationship between maize production values at consecutive time periods. Beyond lag 1, the partial autocorrelation coefficients rapidly declined and fluctuated around zero, with no significant partial autocorrelations observed at higher lags. This suggests that the immediate past value (lag 1) had the most substantial direct influence on the current maize production value, while the influence of more distant lags is negligible when the effect of lag 1 is accounted for. The presence of significant autocorrelations at multiple lags indicated serial dependence in the maize production data. The results were similar even after testing for autocorrelation on squared maize production data suggesting the need for making adjustments to the series using detrending.

Detrending was applied to the maize production time series. Thereafter, ACF and PCF plots were used to verify the removal of autocorrelation from the series. The ACF and PACF plots for the detrended maize production series validates the effectiveness of detrending in addressing autocorrelation for this dataset. ACF and PACF plots were generated to fully confirm that the original and squared values of both precipitation and temperature do not exhibit significant autocorrelation.

4.5 Stationarity, skewness, kurtosis and normality tests

This section presents stationarity test results for detrended maize production, precipitation and temperature using KPSS, PP and ADF tests, skewness and kurtosis results as well as normality test results using Shapiro tests, QQ and density plots.

Table 4.10: Test for stationarity for maize production using KPSS, PP and ADF tests

Test	Test statistic	C.V (10pct)	C.V (5pct)	C.V (1pct)	Conclusion
KPSS	0.1202	0.119	0.146	0.216	Stationary
PP	−4.0844	−12.9	−7.7	−5.5	Non-stationary
ADF	−3.0112	−3.18	−3.50	−4.15	Non-stationary

Note: C.V means critical values

Table 4.11: Test for stationarity for detrended maize production using KPSS, PP and ADF tests

Test	Test statistic	C.V (10pct)	C.V (5pct)	C.V (1pct)	Conclusion
KPSS	0.1202	0.119	0.146	0.216	Stationary
PP	−17.6153	−12.9	−7.7	−5.5	Stationary
ADF	−3.0112	−3.18	−3.50	−4.15	Non-stationary

Table 4.12: Test for stationarity for temperature using KPSS, PP and ADF tests

Test	Test statistic	C.V (10pct)	C.V (5pct)	C.V (1pct)	Conclusion
KPSS	0.0476	0.119	0.146	0.216	Stationary
PP	−34.3494	−12.9	−7.7	−5.5	Stationary
ADF	−4.1749	−3.18	−3.50	−4.15	Stationary

Table 4.13: Test for stationarity for precipitation using KPSS, PP and ADF tests

Test	Test statistic	C.V (10pct)	C.V (5pct)	C.V (1pct)	Conclusion
KPSS	0.0814	0.119	0.146	0.216	Stationary
PP	-37.722	-12.9	-7.7	-5.5	Stationary
ADF	−4.4654	−3.18	−3.50	−4.15	Stationary

Non-stationarity implies that statistical properties of the time series, such as mean, variance, and covariance, change over time. This can lead to a changing dependence structure, which complicates the modelling with a single copula. If the data exhibits trends or seasonality, the dependence captured by the copula might not be representative of the underlying relationship between variables. As a result, testing for stationarity is paramount before any modeling. This was done using the KPSS, PP and ADF tests as shown above.

Maize production is stationary at all levels of significance when tested using the KPSS. However, the series appears non-stationary when tested using PP and ADF. When the maize series is detrended, both the KPSS and PP tests indicate stationarity yet ADF still points to the series being non-stationarity. This study assumes that the non-stationarity indicated by the ADF test is insignificant to affect the copula modelling. The assumption is predicated on the smooth density plots for the detrended maize production series. For the other two variables (temperature and precipitation), the KPSS, PP and ADF test statistics indicates that the null hypothesis of stationarity cannot be rejected, meaning the temperature and precipitation series are stationary.

Table 4.14: Skewness and kurtosis of maize production, detrended maize production, precipitation and temperature

Variable	Skewness	Kurtosis
Maize production	0.2699161	1.687029
Detrended maize production	0.5140849	3.184093
Temperature	0.5625599	3.146388
Precipitation	-0.5431428	3.98702

Maize production had a positive skewness of 0.269 indicating a mild right-skew in the data, suggesting more frequent values below the mean. A kurtosis of 1.687 indicated a leptokurtic distribution with heavier tails and more outliers. On the other hand, detrended maize production also had a positive skewness of 0.514 indicating a moderate positive skew. This means that the distribution is asymmetrical with a longer tail on the right side, suggesting that there are relatively more high-value outliers compared to a perfectly symmetric distribution. Most values still cluster around the center, but the tail is noticeably extended to the right.

Temperature had a moderate positive skewness of 0.563 pointing to a longer right tail which indicated some higher temperatures are more spread out. The kurtosis (3.146) was close to 3, suggesting a distribution similar to normal (mesokurtic), with moderate tail weight. Precipitation had a negative skewness of -0.543 implied a left skew, indicating more frequent higher values. The kurtosis of 3.987 suggested a distribution with heavier tails (leptokurtic) than normal, indicating a higher likelihood of extreme precipitation events. Lastly, the three variables under study were fitted to the normal, exponential, log-normal and gamma univariate distributions to visualize the best fit. The normal distribution appeared to outpace the rest for the three variables.

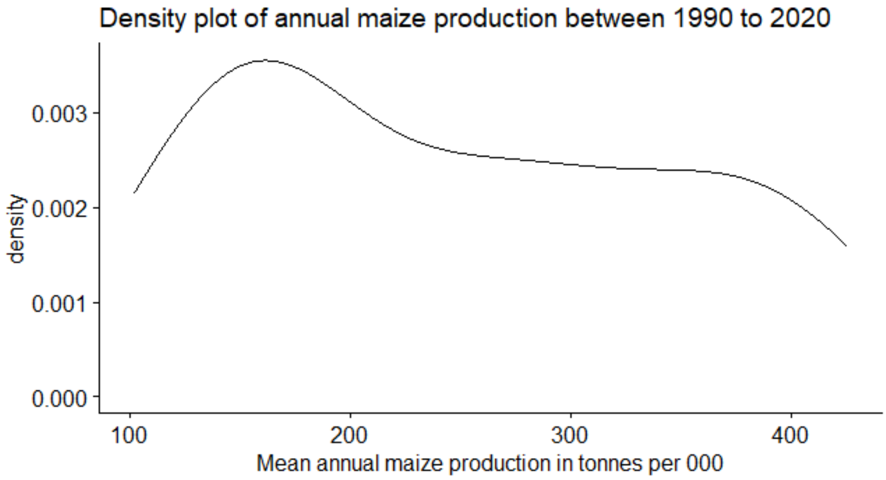


Figure 4.25: Density plot of annual maize production between 1990 to 2020

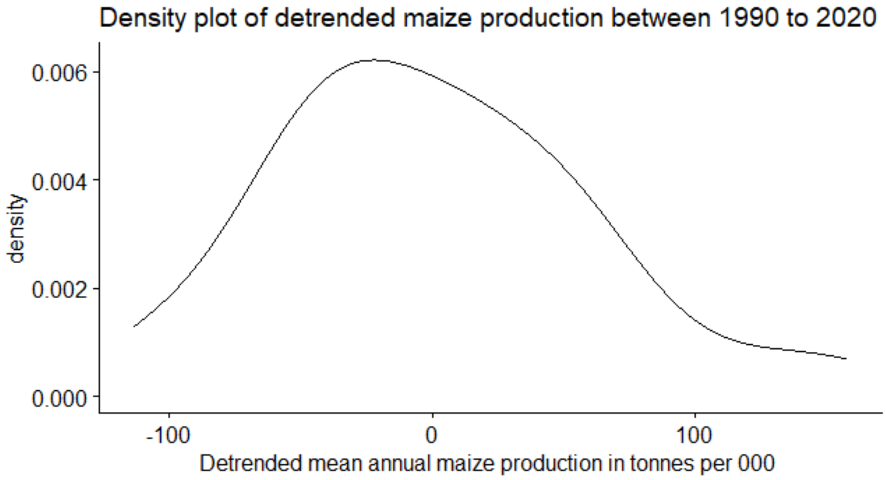


Figure 4.26: Density plot of detrended maize production between 1990 to 2020

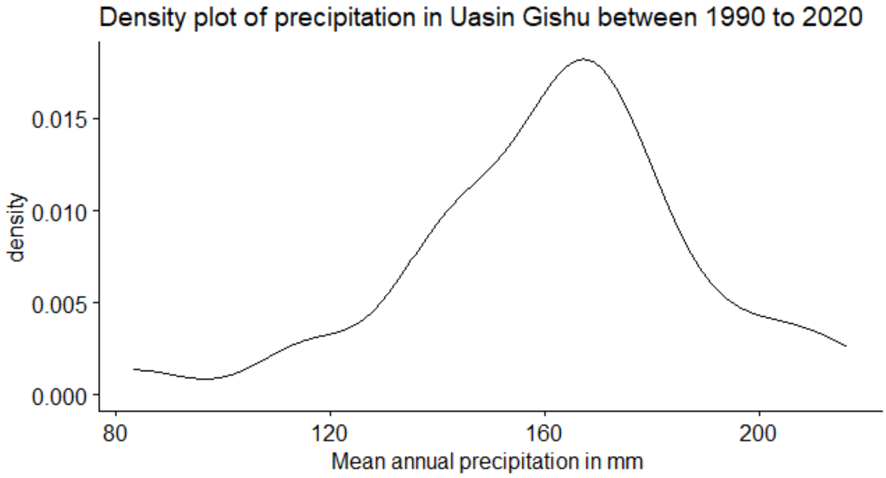


Figure 4.27: Density plot of annual mean precipitation between 1990 to 2020

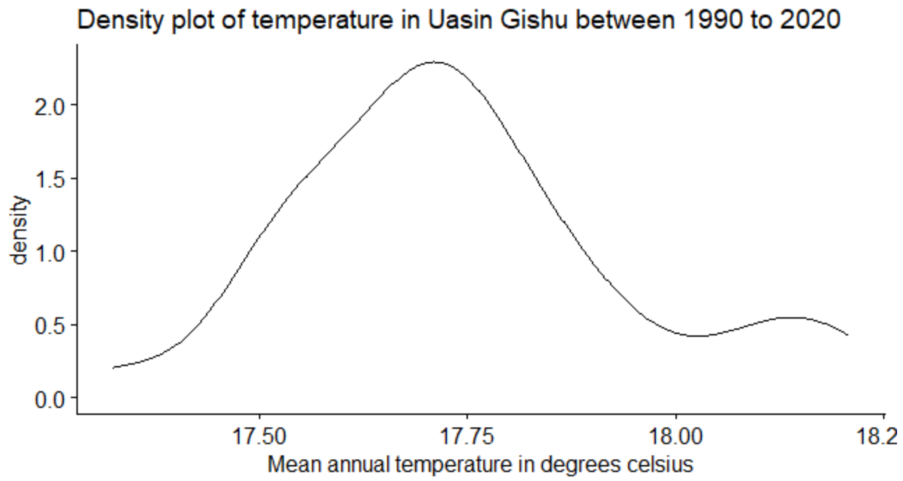


Figure 4.28: Density plot of annual mean temperature between 1990 to 2020

Normality is a simplifying assumption used when dealing with various datasets. The maize production, detrended maize production, temperature and precipitation were tested for normality using common three methods: the density plots, QQ plots and Shapiro test. A density plot is a graphical representation used to visualize the distribution of a continuous variable. It is a smoothed version of a histogram and shows the probability density of the data at different values.

The detrended maize production and temperature plots are positively skewed while the precipitation plot has the tails skewed to the left. The implications of a left-skewed distribution is that the mean is typically less than the median because the lower values pull the mean down more than the median. On other hand, the implications of a right-skewed distribution, the mean is typically greater than the median because the higher values pull the mean up more than the median.

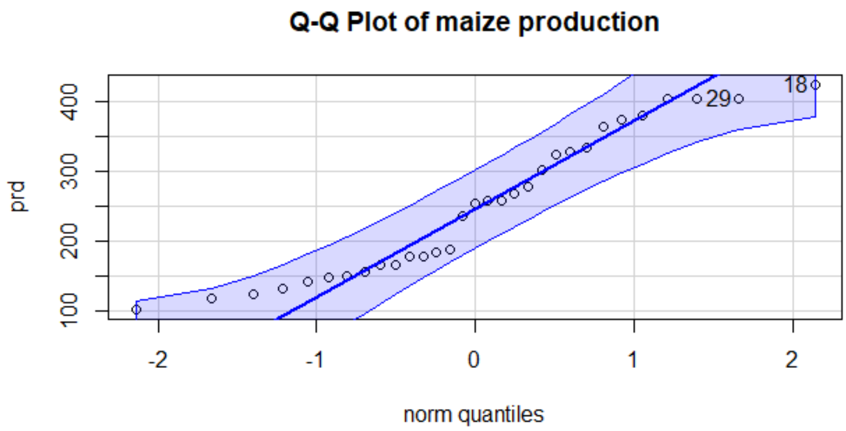


Figure 4.29: QQ plot for maize production

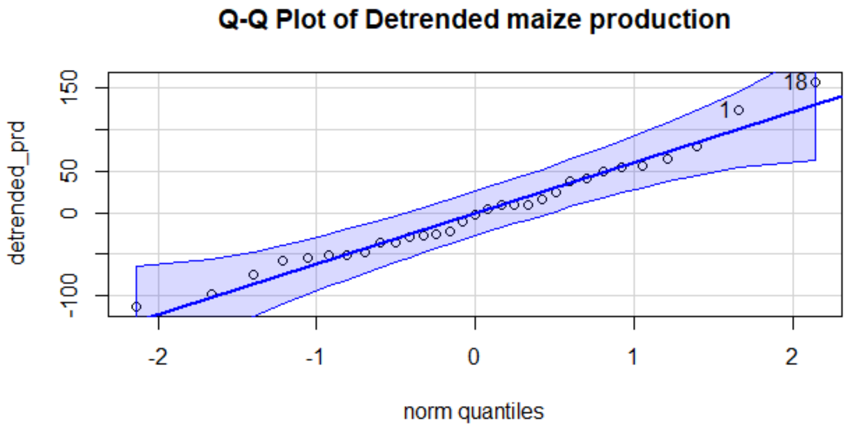


Figure 4.30: QQ plot for detrended maize production

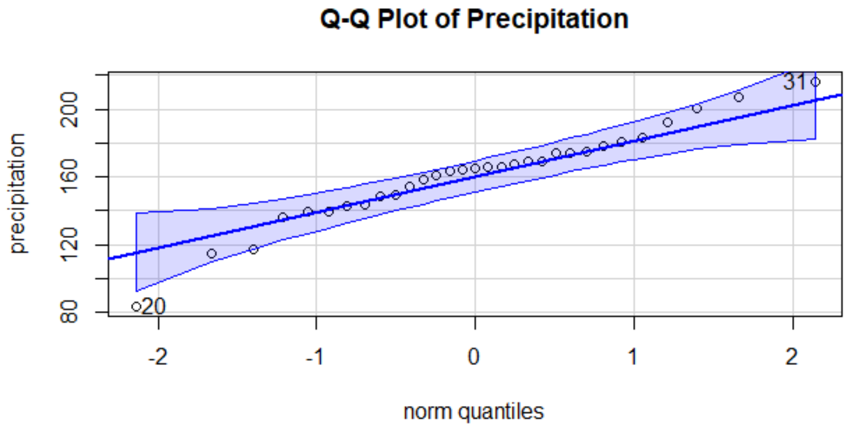


Figure 4.31: QQ plot for precipitation

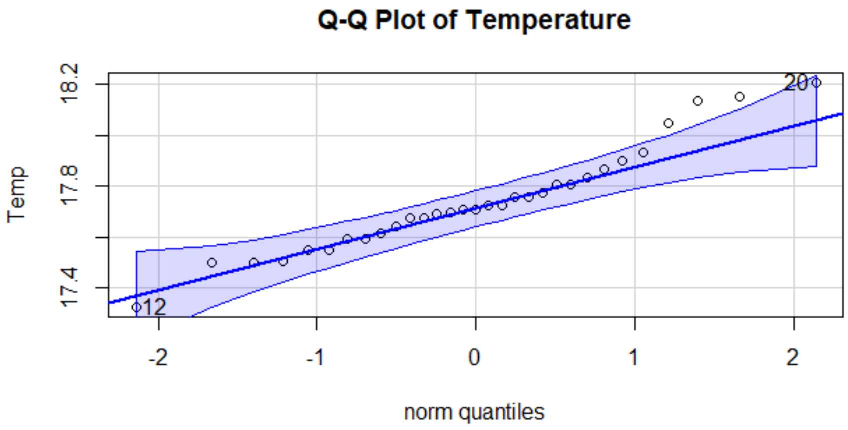


Figure 4.32: QQ plot for temperature

A Quantile-Quantile (QQ) plot is a graphical tool to help assess if a dataset follows a particular distribution, usually a normal distribution. If the points on the QQ plot form an approximate straight line, it suggests that the data follows the specified distribution while any deviations from line point to skewness depending on the direction. The Shapiro test was employed to provide definitive proof of normality in the said datasets.

Table 4.15: Shapiro test for maize production, detrended maize production, precipitation and temperature

Variable	Test statistic	p-value
Maize production	0.91991	0.02348
Detrended maize production	0.97557	0.6819
Temperature	0.95585	0.2258
Precipitation	0.96407	0.3722

The Shapiro-Wilk normality test was conducted on three variables: maize production, temperature, precipitation. The results indicate that original maize production data deviates from a normal distribution, as evidenced by a W statistic of 0.91991 and a p-value of 0.02348, which is less than the conventional significance level of 0.05. This suggests that the maize production data may not be normally distributed. However, detrending resolved non-normality as indicated by the p-value of 0.6819. In contrast, temperature and precipitation data do not show significant deviation from normality, with p-values of 0.2258 and 0.3722, respectively. These results suggest that both temperature and precipitation data are approximately normally distributed, allowing for the use of parametric statistical techniques that assume normality in subsequent analyses.

4.6 Bivariate Copula fitting by MLE and iTau

This section presents the bivariate copula fitting results for detrended maize production, precipitation and temperature using MLE and iTau methods.

Table 4.16: Bivariate Copula fitting results for temperature and precipitation using both the MLE and iTau methods

Method	Copula type	Parameters	Kendall's tau	logLik	AIC	BIC
MLE	Gaussian	par: -0.59	-0.4	5.1	-8.21	-6.77
	Student t	par: -0.58, par2: 3.53	-0.39	5.71	-7.42	4.56
	Rotated Clayton 90°	par: -1.13	-0.36	5.06	-8.11	-6.68
	Rotated Gumbel 270°	par: -1.67	-0.4	5.68	-9.35	-7.92
	Frank	par -4.03	-0.39	5.08	-8.15	-6.57
iTau	Gaussian	par: -0.56	-0.38	5.08	-8.15	-6.72
	Student t	par: -0.56, par2: 3.5	-0.38	5.71	-7.41	4.54
	Rotated Clayton 90°	par: -1.22	-0.38	5.03	-8.05	-6.62
	Rotated Gumbel 270°	par: -1.61	-0.38	5.65	-9.31	-7.87
	Frank	par -3.89	-0.38	4.99	-7.99	-6.56

Copula fitting involves modelling the dependency structure between multiple random variables using a copula. A copula is a statistical tool that allows you to separate the marginal distributions of individual variables from their joint dependency structure. Dependence was tested between temperature vs precipitation, detrended maize production vs precipitation and detrended maize production vs temperature using the MLE method and i-tau method as shown above.

The dependence structure between temperature and precipitation using multiple copula families was evaluated by fitting through maximum likelihood estimation (MLE) and indirect tau (iTau) methods. The analysis aimed to select the copula family that best models the joint distribution, comparing goodness-of-fit measures across Gaussian, t, Rotated Clayton, Rotated Gumbel, and Frank copulas. For each copula family, parameter estimation was performed using both the MLE and iTau methods, and the best fit was selected based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Kendall's tau was computed to assess dependence strength, and upper and lower tail dependencies (TD) were noted.

Among the evaluated families, the rotated Gumbel (270°) copula provided the best fit according to AIC (-9.35) and BIC (-7.92). This copula reflected a moderate negative dependence with Kendall's tau of -0.40, suggesting a consistent, though modest, inverse relationship between temperature and precipitation. The rotated Gumbel (270°) copula also yields the best fit using the iTau method, with an AIC of -9.31 and BIC of -7.87. Across the MLE and iTau methods, the rotated Gumbel (270°) copula emerged as the best-fitting model for the dependence between temperature and precipitation. This copula's fit statistics and negative Kendall's tau suggest a moderate inverse relationship, with a consistent structure across estimation methods.

Table 4.17: Bivariate Copula fitting results for detrended maize production and precipitation using both the MLE and iTau methods

Method	Copula type	Parameters	Kendall's tau	logLik	AIC	BIC
MLE	Gaussian	par: 0.29	0.19	0.96	0.08	1.52
	Student t	par: 0.3, par2: 30	0.19	0.96	2.08	4.95
	Survival Clayton	par: 0.37	0.16	0.63	0.73	2.17
	Gumbel	par: 1.23	0.18	0.93	0.15	1.58
	Frank	par 1.66	0.18	1.04	−0.08	1.35
iTau	Gaussian	par: 0.29	0.19	0.96	0.09	1.52
	Student t	par: 0.29, par2: 9.66	0.19	0.92	2.16	5.03
	Survival Clayton	par: 0.46	0.19	0.6	0.8	2.24
	Gumbel	par: 1.23	0.19	0.92	0.15	1.58
	Frank	par 1.74	0.19	1.04	−0.08	1.36

Next, this study assessed the copula selection and fitting results for the relationship between detrended maize production and precipitation. Based on AIC and BIC, the Gaussian copula yield the best fit. It demonstrated a moderate positive dependence with a Kendall's tau of 0.19, AIC of 0.08 and a BIC of 1.52. The Gaussian copula was also the best fit under the iTau method, with a Kendall's tau of 0.19, AIC of

0.09, and BIC of 1.52, suggesting a weak positive dependence. Across the MLE and iTau methods, the Gaussian copula showed the most consistent performance, followed closely by the Frank copula, both having relatively low AIC values.

Table 4.18: Bivariate Copula fitting results for detrended maize production and temperature using both the MLE and iTau methods

Method	Copula type	Parameters	Kendall's tau	logLik	AIC	BIC
MLE	Gaussian	par: -0.18	-0.12	0.35	1.3	2.73
	Student t	par: -0.18, par2: 30	-0.11	0.25	3.51	6.37
	Rotated Clayton 90°	par: -0.16	-0.08	0.15	1.71	3.14
	Rotated Gumbel 270°	par: -1.09	-0.08	0.16	1.68	3.12
	Frank	par -0.84	-0.09	0.28	1.45	2.88
iTau	Gaussian	par: -0.18	-0.12	0.35	1.3	2.73
	Student t	par: -0.18, par2: 9.41	-0.12	0.0	4.0	6.87
	Rotated Clayton 90°	par: -0.26	-0.12	0.09	1.81	3.24
	Rotated Gumbel 270°	par: -1.13	-0.12	0.12	1.76	3.2
	Frank	par -1.06	-0.12	0.26	1.49	2.92

Lastly, the dependency structure between detrended maize production and temperature was analyzed as follows: The MLE method was applied to each copula family, and selection criteria were determined using AIC (Akaike Information Criterion). The AIC and BIC indicated that Gaussian copula (AIC = 1.3, BIC = 2.73) provided the best fit compared to other copulas. The negative Kendall's Tau across models suggested independence between detrended maize production and temperature. Using the Inverse Kendall's Tau method, which focuses on rank-based dependence estimation, results for each copula family were computed. Consistent with MLE, the iTau method suggests the Gaussian copula (AIC = 1.3, BIC = 2.73) offers a slightly better fit than others.

4.7 Trivariate Copula fitting by MLE and iTau

This section presents the trivariate copula fitting results for detrended maize production, precipitation and temperature using MLE and iTau methods.

Table 4.19: Trivariate Copula fitting results for detrended maize production and temperature using both the MLE and iTau methods

Method	Copula type	Parameters	logLik	AIC	BIC
MLE	Gaussian	rho 1:0.297, rho 2:-0.201, rho 3:-0.588	6.074	-6.147	-1.845
	Student t	rho 1: 0.296, rho 2:-0.176, rho 3:-0.590 df: 7.702	6.276	-4.553	1.183
	Clayton	α: 0	0	2	3.434
	Gumbel	α: 1	0	2	3.434
	Frank	α: 0	0	2	3.434
iTau	Gaussian	rho 1:0.290, rho 2:-0.182, rho 3:-0.562	–	–	–
	Student t	rho 1: 0.290, rho 2:-0.182, rho 3:-0.562 df: 7.702	–	–	–
	Clayton	α: 0.1	–	–	–
	Gumbel	α: 1.077	–	–	–
	Frank	α: -1.072	–	–	–

In this study, a three-dimensional dataset with variables detrended maize production, precipitation and temperature was fitted to parametric copula models. Different copula families (Gaussian, Student-t, Clayton, Gumbel, Frank) were compared using parameter estimation methods such as maximum likelihood and inversion of Kendall's tau. The most suitable model was selected using AIC and BIC.

The normal copula yielded pair copula estimates which describe the dependence between detrended maize production and precipitation, maize production and temperature as well as precipitation and temperature. The parameter estimates and their corresponding standard errors were aligned closely for both the maximum likelihood and itau methods suggesting consistency in dependence patterns.

The student-t copula model fitted using maximum likelihood method did not yield any standard errors for the estimated parameters. This indicated potential issues with estimation precision due to the small sample size (n = 31) as well as the numerical instability of the estimation method used. The degrees of freedom (df = 7.702) suggested the dependence structure was close to Gaussian. Equally, the student-t copula model fitted using itau method assumed fixed degrees of freedom thus limiting flexibility in capturing tail dependence.

The clayton copula model fitted using the maximum likelihood method had an alpha equal to zero which suggests independence among the variables under study. In addition, the log likelihood was also zero with no corresponding standard errors which was most likely due to insufficient data (n = 31) and pointing to a poor fit or an unreliable boundary parameter estimate. The parameter estimate had no corresponding standard error suggesting optimization failure at the parameter boundary. The low AIC and BIC values point to a poor model fit for the Clayton trivariate copula. On the other hand, the alpha value generated from the itau method was a negative value which is outside the theoretical range required by the Clayton copula. This result suggested that the Clayton copula is not a good fit for the data under study.

Similarly, the Gumbel copula model fitted using the maximum likelihood method had an alpha value equal to one which reduces it to the independence copula implying no tail dependence. The log likelihood of zero and lack of a corresponding standard error for the parameter estimate indicated a poor fit or an unreliable boundary parameter estimate, just as in the Clayton copula model. The Gumbel model fitted using the itau model yielded an alpha value of 1.091 hence pointing to very weak upper tail dependence among the three variables under study. This method encountered negative Kendall's tau values, which were replaced with 0 to ensure alpha is greater than or equal to one. This may indicate weak or inconsistent dependence in the data. While the parameter estimates for the Frank copula model fitted using the maximum likelihood method yielded similar issues as with the Clayton and Gumbel copulas, the alpha value estimated using the itau method was -0.581 implying negative dependence for the variables under study.

In conclusion, the Archimedean copula models (Clayton, Gumbel, Frank) did not display a good fit for the data and were susceptible to estimation errors. This limited the interpretability and reliability of their parameter estimates. Only the student t and gaussian copula models displayed plausible results and were most appropriate for capturing the dependencies between detrended maize production, precipitation, and temperature. In particular, the trivariate gaussian copula outpaced the t copula when their AIC and BIC values were compared.

4.8 Conditional copulas fitting by MLE and iTau

This section presents the conditional copula fitting results for detrended maize production, precipitation and temperature using MLE and iTau methods.

Table 4.20: Pair copula fitting estimations for detrended maize production, precipitation and temperature

Pair copula	Copula type	Parameters	Kendall's tau	p-value	logLik	AIC	BIC
V1|V2	Gumbel	1.23	0.18	0.14	0.93	0.15	1.58
V1|V3	Rotated Gumbel 270°	−1.09	−0.08	0.36	0.16	1.68	3.12
V2|V3	Rotated Gumbel 90°	−1.67	−0.4	<0.01	5.68	−9.35	−7.92

Note: V1, V2 and V3 refer to detrended maize production, precipitation and temperature respectively

Table 4.21: Conditional copulas estimations for detrended maize production, precipitation and temperature using MLE and iTau

Conditional copula	Method	Type	Parameters	Kendall's tau	p-value	logLik	AIC	BIC
V1,V3|V2	MLE	Independence	0	0	0.72	0	0	0
	iTau	Independence	0	0	0.72	0	0	0
V2,V3|V1	MLE	Rotated Gumbel 90°	−1.64	−0.39	<0.01	5.51	−9.03	−7.59
	iTau	Rotated Gumbel 90°	−1.52	−0.34	<0.01	5.38	−8.76	−7.32
V1,V2|V3	MLE	Gaussian	0.84	0.64	<0.01	17.63	−33.26	−31.83
	iTau	Gaussian	0.82	0.61	<0.01	17.5	−33	−31.57

Note: V1, V2 and V3 refer to detrended maize production, precipitation and temperature respectively

4.9 Vine copula fitting

This section presents the vine copula fitting results for detrended maize production, precipitation and temperature using parametric and non-parametric methods.

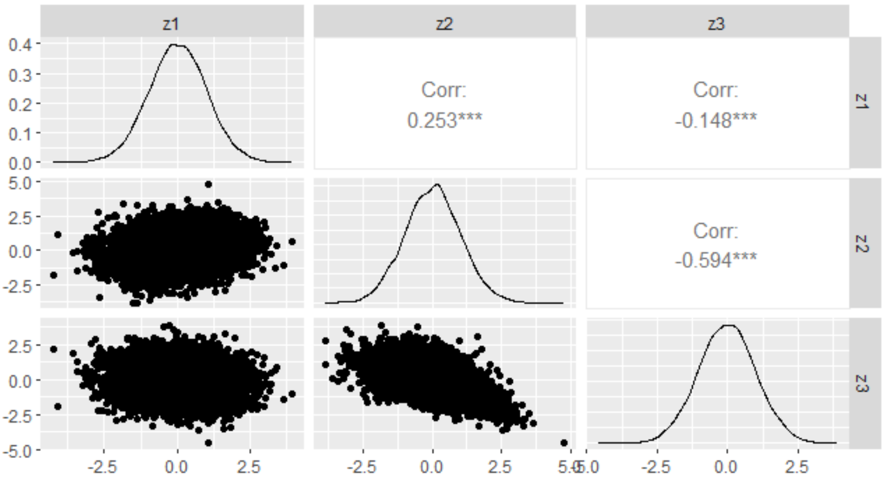


Figure 4.33: Pair plot showing histograms, scatter plots and correlation coefficients for simulated data for detrended maize production and temperature given precipitation

Table 4.22: Vine copula fitting results for detrended maize production and temperature given precipitation

Copula family	Conditioned	Conditioning	Copula type	Parameters	Kendall's tau	logLik
Elliptical	1,2	-	Student-t	0.26, 45.93	0.166	340
	2,3	-	Student-t	−0.6, 8.5	−0.406	2200
	1,3	2	Gaussian	0.00071	0.00045	0.0025
Archimedean	1,2	-	Frank	1.7	0.1805	370
	2,3	-	Rotated Gumbel 270°	−1.67	−0.407	2500
	1,3	2	Frank	0.02	0.0022	0.055
Non-parametric	1,2	-	tll	[30*30 grid]	0.1734	441
	2,3	-	tll	[30*30 grid]	−0.3957	2521
	1,3	2	tll	[30*30 grid]	0.0024	64

Table 4.23: Comparison of vine copula model fits for detrended maize production and temperature given precipitation

Model type	logLik	No. of parameters	AIC	BIC
Elliptical	2578.21	4	−5148.42	−5119.57
Archimedean	2834.78	3	−5663.57	−5641.93
Non-parametric	3026.4	397.99	−5256.82	−2387.21

Note: Elliptical refers to Gaussian and Student t copulas while Archimedean refers to Clayton, Gumbel and Frank copulas

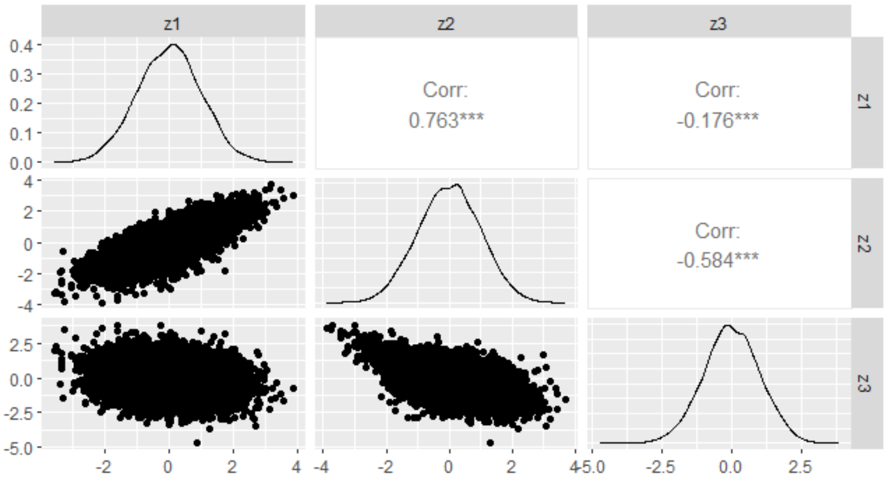


Figure 4.34: Pair plot showing histograms, scatter plots and correlation coefficients for simulated data for detrended maize production and precipitation given temperature

Table 4.24: Vine copula fitting results for detrended maize production and temperature given precipitation

Copula family	Conditioned	Conditioning	Copula type	Parameters	Kendall's tau	logLik
Elliptical	1,3	-	Gaussian	-0.18	-0.11	157
	3,2	-	Student-t	-0.58, 7.97	-0.4	2162
	1,2	3	Gaussian	0.83	0.62	5849
Archimedean	1,3	-	Frank	-1	-0.11	135
	3,2	-	Rotated Gumbel 270°	-1.7	-0.40	2330
	1,2	3	Rotated Gumbel 180°	2.5	0.61	5709
Non-parametric	1,3	-	tll	[30*30 grid]	-0.11	228
	3,2	-	tll	[30*30 grid]	-0.39	2390
	1,2	3	tll	[30*30 grid]	0.62	6165

Table 4.25: Comparison of vine copula model fits for detrended maize production and precipitation given temperature

Copula family	logLik	No. of parameters	AIC	BIC
Elliptical	8166.82	4	-16325.64	-16296.80
Archimedean	8174.48	3	-16342.96	-16321.33
Non-parametric	8782.28	402.43	-16759.69	-13858.02

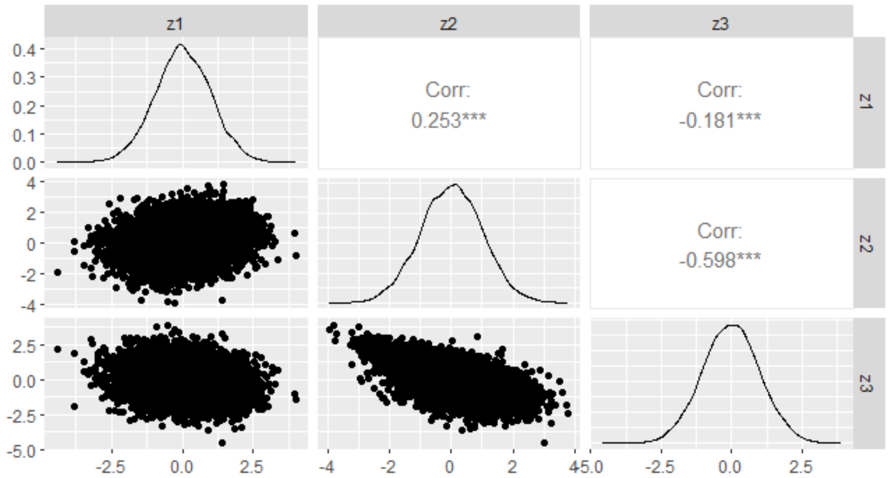


Figure 4.35: Pair plot showing histograms, scatter plots and correlation coefficients for simulated data for precipitation and temperature given detrended maize production

Table 4.26: Vine copula fitting results for precipitation and temperature given detrended maize production

Copula family	Conditioned	Conditioning	Copula type	Parameters	Kendall's tau	logLik
Elliptical	2,1	-	Gaussian	0.25	0.16	334
	1,3	-	Gaussian	−0.18, 7.97	−0.12	166
	2,3	1	Student-t	−0.58, 8.56	−0.39	2113
Archimedean	2,1	-	Frank	1.7	0.18	365
	1,3	-	Frank	−1.1	−0.12	152
	2,3	1	Rotated Gumbel 90°	1.6	−0.39	2276
Non-parametric	2,1	-	tll	[30*30 grid]	0.17	441
	1,3	-	tll	[30*30 grid]	−0.11	249
	2,3	1	tll	[30*30 grid]	−0.38	2332

Table 4.27: Comparison of vine copula model fits for precipitation and temperature given detrended maize production

Copula family	logLik	No. of parameters	AIC	BIC
Elliptical	2613.22	4	−5218.44	−5189.60
Archimedean	2794.30	3	−5582.59	−5560.96
Non-parametric	3021.22	393.72	−5255	−2416.12

One of the main limitations of using the trivariate copulas was that the sample size, n = 31, was very small to generate stable parameter estimates, especially in three dimension. In addition, the estimation errors encountered when using Archimedean copulas limited the interpretability and reliability of parameters obtained. To counter these downsides, a more advanced analysis was required namely vine copulas. Vine copulas are a flexible class of copulas used to model complex dependence structures in multivariate data. They decompose multivariate distributions into a series of conditional bivariate copulas, organized in a hierarchical structure called a vine. They also allow for the modelling of intricate relationships while maintaining computational efficiency. Moreso, vine copulas are known for their ability to capture asymmetric and tail dependencies.

In this study, modelling using vine copulas yielded three possible pair copula structures namely: detrended maize production and temperature given precipitation, precipitation and temperature given detrended maize production as well as detrended maize production and precipitation given temperature. Each structure assumed a different conditional dependence scheme, allowing the examination of the relationships between the variables. After specifying the pair copulas and a vine structure, the vine copula models were fitted using the suitable parameters obtained from the bivariate copula analysis under section 4.6 above. 10,000 random samples were generated from a vine copula model and transformed to standard normal margins. Thereafter, empirical marginal distributions were used to estimate pseudo-observations from the generated data. The generated pseudo-observations were fitted into vine copula models using maximum likelihood estimation to obtain the conditional copulas. The copula families used include gaussian, Student-t, Clayton, Gumbel and Frank.

For the three pair copula structures, nonparametric grid-based copulas provided better log-likelihood values, indicating a more flexible fit to the data when compared to gaussian and Archimedean copulas. This is because nonparametric copulas can model complex dependence structures, including non-elliptical, asymmetric, and tail dependencies, without assuming a specific functional form. They also capture arbitrary shapes of the dependence structure, making it suitable for data with irregular patterns, in this case, the detrended maize production dataset. However, they do not require assumptions about the underlying distribution, hence prone to misspecification. Moreover, nonparametric copulas are prone to overfitting especially when small datasets are used. Therefore, these copulas need a large amount of data to accurately estimate the copula, particularly in higher dimensions. With sparse data, the estimates may be unreliable. Lastly, nonparametric copulas are less interpretable than parametric copulas, as there is no closed-form expression or parameters to describe the dependence structure. Nevertheless, the tau estimates for both parametric and non-parametric copulas were not significantly different hence both estimates can be relied upon. In general, non-Archimedean copulas (Gaussian and t) provided a better fit for the data compared to the Archimedean and non-parametric copulas as evidenced by their low AIC and BIC values.

Conditioning precipitation on maize production and temperature using both the elliptical and Archimedean copulas indicated that detrended maize production and temperature have little dependence as signified by the very negligible tau values. In other words, the residual dependence between maize production and temperature after accounting for the influence of precipitation points to an almost independent relationship between the two variables. This finding further confirms the veracity of the weak positive dependence (Kendall's tau of -0.12) which was first highlighted by the bivariate gaussian copula. Similarly, conditioning temperature on maize production and precipitation caused the Kendall's tau between the two variables to triple from 0.18, as indicated by the bivariate Frank copula, to 0.62. This indicates a very strong dependence between maize production and precipitation as empirically evidenced. Interestingly, conditioning maize production on precipitation and temperature had no effect on the dependence between the variables with the tau remaining constant at -0.39, just as it was in the bivariate case. In conclusion, such complex dependence structures were only visible due to the use of vine copulas and points to the need to a careful analysis of relationships between variables by policy makers using advanced mathematical tools.

In summary, by isolating these dependencies, stakeholders gain insights into how weather variables interact to impact maize production. Conditional copulas reveal how combinations of weather extremes (e.g., high temperature + low rainfall) affect yields. As a result, farmers can prioritize adaptive strategies by investing in drought-resistant seeds if heat-dryness interactions are critical or by adjusting planting schedules to avoid peak heat periods coinciding with low rainfall. On the other hand, policymakers can target subsidies for irrigation infrastructure in high-risk zones and also develop early-warning systems for extreme weather combinations. Lastly, risk managers can design insurance products or weather derivatives triggered by specific weather thresholds (e.g., payout if rainfall < 400mm and temperature > 32°C). Financial institutions can also adjust loan terms based on localized risk profiles.

Chapter 5: Conclusion and Recommendations

5.1 Conclusion

This chapter presents a comprehensive analysis of the complex relationships between maize production and key climatic variables - precipitation and temperature using advanced copula methods. The study was motivated by the need to understand how these weather factors interact to influence agricultural output in a semi-arid region, particularly given the increasing climate variability observed over the 30-year study period (1990-2020). The analysis began with a thorough examination of the dataset, which revealed significant trends: maize cultivation area expanded by 126%, yields more than doubled from 1.94 tonnes/ha to 4.96 tonnes/ha, and distinct weather extremes like the 1997 El Niño and 2009 drought caused notable production declines. These patterns underscored the importance of developing robust models to capture climate-agriculture interactions. Further the data pre-treatment phase established a robust foundation for subsequent modeling. Detrending successfully addressed autocorrelation in maize production (Ljung-Box p-value improved from 0.000003 to 0.02936), while stationarity tests (KPSS, PP, ADF) confirmed the suitability of processed data for copula applications. Normality checks revealed maize production's initial deviation from Gaussian distribution (Shapiro-Wilk p = 0.02348), resolved post-detrending, enabling reliable parametric modeling.

Methodologically, the study employed a sophisticated copula framework after addressing critical data preprocessing steps. Initial tests revealed strong autocorrelation in maize production data, which was successfully mitigated through detrending. Stationarity tests using KPSS, PP, and ADF methods confirmed the suitability of the processed data for copula modeling. The analysis progressed through three key stages: bivariate copulas to examine pairwise relationships, trivariate copulas for joint dependencies, and vine copulas to unravel conditional structures. Among bivariate models, a rotated Gumbel copula best captured the moderate negative dependence between temperature and precipitation (Kendall's $\tau = -0.40$), while Gaussian copulas were optimal for maize-weather relationships. These results highlighted precipitation's dual role—beneficial up to 161 mm/month but detrimental beyond 561 mm—and temperature's subtle threshold effects, masked in raw data but revealed through detrending.

The trivariate analysis faced challenges with Archimedean copulas due to sample size limitations, but Gaussian and Student-t copulas provided reliable estimates of the three-way dependencies. While Archimedean copulas (Clayton, Gumbel) failed to converge reliably, elliptical copulas—particularly Gaussian—provided stable estimates of joint dependencies (AIC = 1.3, BIC = 2.73). The Gaussian copula's dominance underscored linear dependencies' persistence in three-dimensional space, though its inability to capture tail risks suggested room for refinement with larger datasets.

Vine copula analysis revolutionized the interpretation of conditional dependencies. Decomposing trivariate relationships into pairwise structures revealed hidden synergies: conditioning on precipitation tripled the maize-temperature dependence (tau from 0.18 to 0.62), exposing how drought amplifies heat stress impacts. Conversely, conditioning on temperature left maize-precipitation dependence unchanged (tu $= -0.39$), emphasizing precipitation's standalone criticality. Nonparametric vine copulas, while flexible, risked overfitting, reaffirming Gaussian/t-copulas as pragmatic choices for interpretable risk modeling.

The results revealed several critical insights. First, while precipitation showed a weak positive correlation with maize yields (Pearson's r = 0.26), extreme rainfall events (>561 mm) proved detrimental, highlighting a nonlinear threshold effect. Temperature exhibited minimal direct impact, but detrending uncovered subtle negative effects ($\tau = -0.12$), suggesting heat stress becomes relevant beyond certain thresholds. Most significantly, vine copula analysis demonstrated how conditioning on third variables dramatically altered dependence structures - the maize-precipitation relationship strengthened considerably (τ increased to 0.62) when accounting for temperature effects. These findings have important practical implications: farmers in high-heat regions should prioritize drought-resistant varieties, policymakers need to target irrigation investments in drought-prone areas, and insurers can develop innovative products based on specific weather thresholds.

The study's innovative use of vine copulas represents a methodological advance in agro-climatic modeling, providing a framework to uncover complex, conditional relationships that traditional linear approaches miss. However, limitations including the modest sample size (n = 31 years) and exclusion of biotic factors suggest directions for future research. Incorporating higher-resolution satellite data and expanding to multi-hazard scenarios would enhance the model's predictive capability. Ultimately, this research demonstrates how advanced statistical tools can transform our understanding of climate-agriculture interactions, offering stakeholders science-based strategies to enhance resilience in maize-dependent economies facing increasing climate variability.

The copula approach developed here provides a versatile framework adaptable to other crops and regions, marking an important step toward more climate-smart agricultural planning. By bridging statistical rigor with agricultural relevance, this research establishes copulas as indispensable tools for climate-resilient farming, offering a replicable framework to navigate the escalating uncertainties of global warming. Future applications to other crops and regions could further democratize data-driven climate adaptation in vulnerable agro-ecosystems.

5.2 Recommendations

Based on the findings of this study, the following recommendations are presented to guide policy makers:

Climate Adaptation Strategies

Developing strategies to mitigate the impact of precipitation variability. The government can incentivise local manufacturers of irrigation equipment through tax breaks and accelerated capital allowances on capital investments made. This would in turn reduce the final cost to farmers thus encouraging the uptake of drip irrigation among other techniques.

Monitoring and Research

Establishing a robust climatic monitoring system to provide timely data and predictions, enabling farmers to plan better. Adverse variability in weather variables particularly precipitation could affect maize yields, impact the nation's food security and alter the country's balance of payment metrics.

Embracing technology

Cloud seeding techniques have been documented as effective in several places around the world. County governments whose regions contribute significantly to the country's breadbasket in terms of maize production should invest in research and development together with partnering with established players with this industry.

5.3 Limitations of the Study

This study was limited to one of the counties in Kenya. It can be replicated in other administrative units within the country to compare the effectiveness of using copulas when analysing the nexus between agricultural productivity and climatic variables as compared to multiple linear regression and correlation analysis.

5.4 Areas for Further Research

This study was limited by time and resources hence several simplifying assumptions were used as indicated in 1. Future research work can improve on this study by:

- Exploring the interactions between precipitation and other agronomic factors such as mechanization, soil type, use of artificial and organic fertilizers etc to better elucidate their combined effects on maize production.
- Modelling a weather derivative product which can be used by maize farmers to hedge against the risk and uncertainty of weather variables.

References

Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, 44(2):182–198.

AGRA (2021). *Africa Agriculture Status Report: a Decade of Action: Building Sustainable and Resilient Food Systems in Africa.* Number 9. Alliance for a Green Revolution in Africa.

Akpalu, W., Hassan, R. M., and Ringler, C. (2008). Climate variability and maize yield in south africa: results from gme and mele methods. international food policy research institute, ifpri discussion paper 0084, december 2008.

Bello, G. H. M. and Maman, M. N. M. (2015). A ricardian analysis of the impact of temperature and rainfall variability on ag-riculture in dosso and maradi regions of niger republic. *Agricultural Sciences*, 6(7):724–733.

Blanc, É. (2011). *The impact of climate change on crop production in Sub-Saharan Africa.* PhD thesis, University of Otago.

Cabas, J., Weersink, A., and Olale, E. (2010). Crop yield response to economic, site and climatic variables. *Climatic change*, 101(3):599–616.

Chen, X., Fan, Y., and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association*, 101(475):1228–1240.

Chisanga, C. B., Moombe, M., and Elijah, P. (2022). Modelling climate change impacts on maize. *CABI Reviews*, (2022).

Czado, C. and Nagler, T. (2022). Vine copula based modeling. *Annual Review of Statistics and Its Application*, 9(1):453–477.

Di Falco, S., Veronesi, M., and Yesuf, M. (2011). Does adaptation to climate change provide food security? a micro-perspective from ethiopia. *American journal of agricultural economics*, 93(3):829–846.

Dwamena, H. A., Tawiah, K., and Akuoko Kodua, A. S. (2022). The effect of rainfall, temperature, and relative humidity on the yield of cassava, yam, and maize in the ashanti region of ghana. *International Journal of Agronomy*, 2022(1):9077383.

FAO (2021). The state of food and agriculture 2021: making agrifood systems more resilient to shocks and stresses.

Jones, P. G. and Thornton, P. K. (2003). The potential impacts of climate change on maize production in africa and latin america in 2055. *Global environmental change*, 13(1):51–59.

Kariuki, G., Njaramba, J., and Ombuki, C. (2018). Climate change and maize yield in kenya: an econometric analysis. proceedings of the 1st annual international conference held on 17th-19th april 2018, machakos university, kenya.

Karthikeyan, L., Chawla, I., and Mishra, A. K. (2020). A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses. *Journal of Hydrology*, 586:124905.

Kimani, A. K. (2017). *Climate Change Impact on Agriculture: Challenges on Maize Production in Uasin Gishu and Trans-nzoia Counties*. PhD thesis, University of Nairobi.

Li, F., Tang, Y., and Wang, H. J. (2019a). Copula-based semiparametric analysis for time series data with detection limits. *Canadian Journal of Statistics*, 47(3):438–454.

Li, Y., Guan, K., Schnitkey, G. D., DeLucia, E., and Peng, B. (2019b). Excessive rainfall leads to maize yield loss of a comparable magnitude to extreme drought in the united states. *Global change biology*, 25(7):2325–2337.

Lobell, D. B. and Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and forest meteorology*, 150(11):1443–1452.

Riaman, Sukono, Supian, S., and Ismail, N. (2022). Mathematical modeling for estimating the risk of rice farmers' losses due to weather changes. *Computation*, 10(8):140.

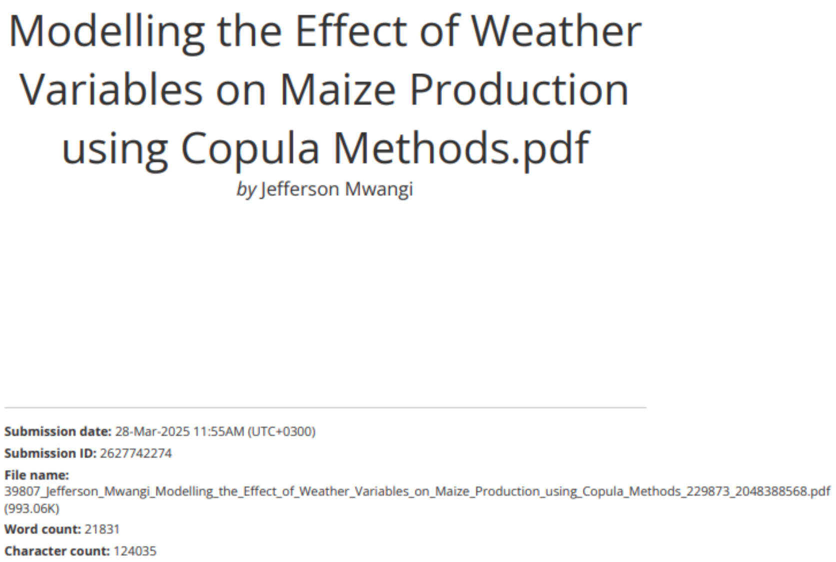
Ribeiro, A. F., Russo, A., Gouveia, C. M., Páscoa, P., and Pires, C. A. (2019). Probabilistic modelling of the dependence between rainfed crops and drought hazard. *Natural Hazards and Earth System Sciences*, 19(12):2795–2809.

Schlenker, W. and Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to us crop yields under climate change. *Proceedings of the National Academy of sciences*, 106(37):15594–15598.

Tschora, H. and Cherubini, F. (2020). Co-benefits and trade-offs of agroforestry for climate change mitigation and other sustainability goals in west africa. *Global Ecology and Conservation*, 22:e00919.

Unnikrishnan, P., Ponnambalam, K., and Karray, F. (2024). Influence of regional temperature anomalies on strawberry yield: A study using multivariate copula analysis. *Sustainability*, 16(9):3523.

Chapter A: Similarity Report



Modelling the Effect of Weather Variables on Maize Production using Copula Methods.pdf

by Jefferson Mwangi

Submission date: 28-Mar-2025 11:55AM (UTC+0300)

Submission ID: 2627742274

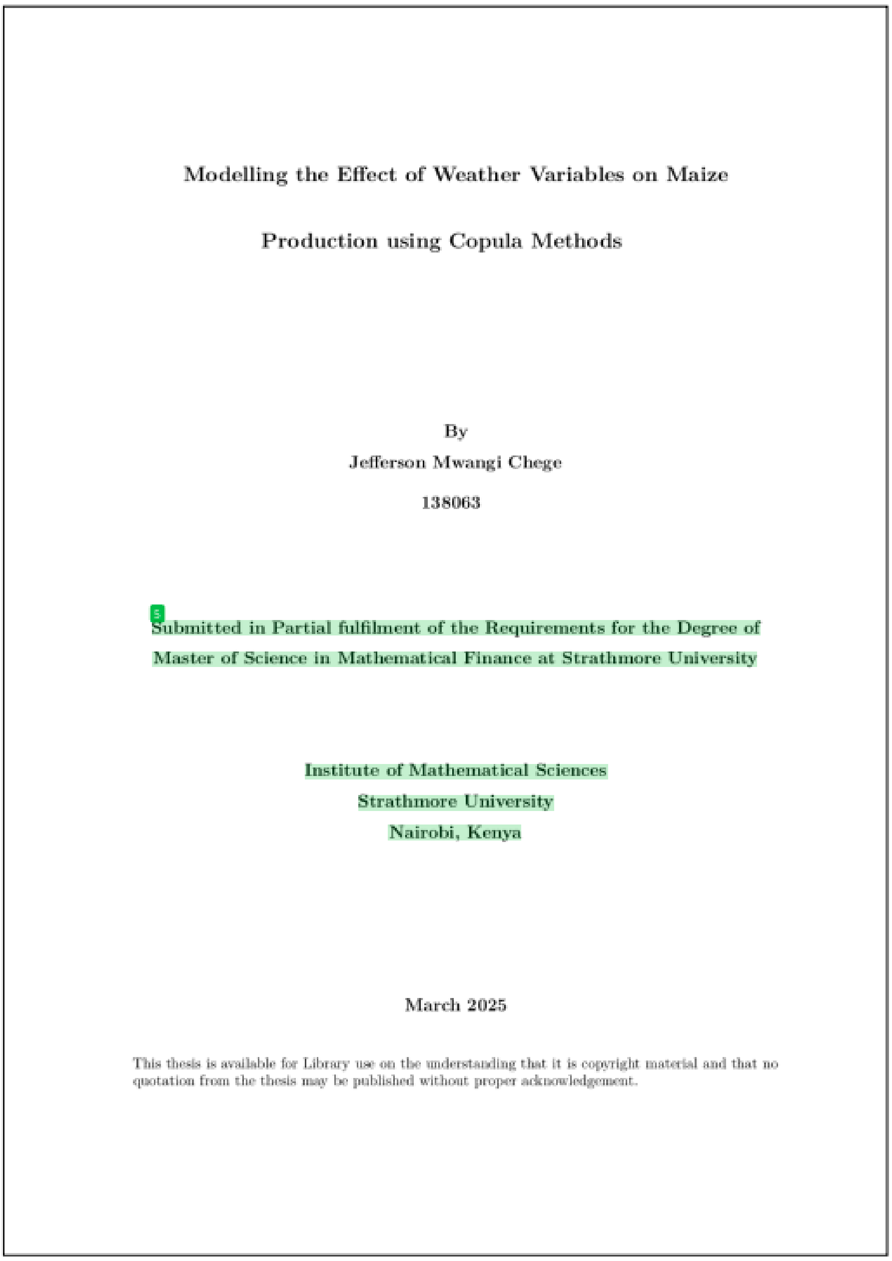
File name:
39807_Jefferson_Mwangi_Modelling_the_Effect_of_Weather_Variables_on_Maize_Production_using_Copula_Methods_229873_2048388568.pdf
(993.06K)

Word count: 21831

Character count: 124035

Figure A.1: Similarity report page 1

Chapter B: Similarity Report



Modelling the Effect of Weather Variables on Maize

Production using Copula Methods

By

Jefferson Mwangi Chege

138063

Submitted in Partial fulfilment of the Requirements for the Degree of
Master of Science in Mathematical Finance at Strathmore University

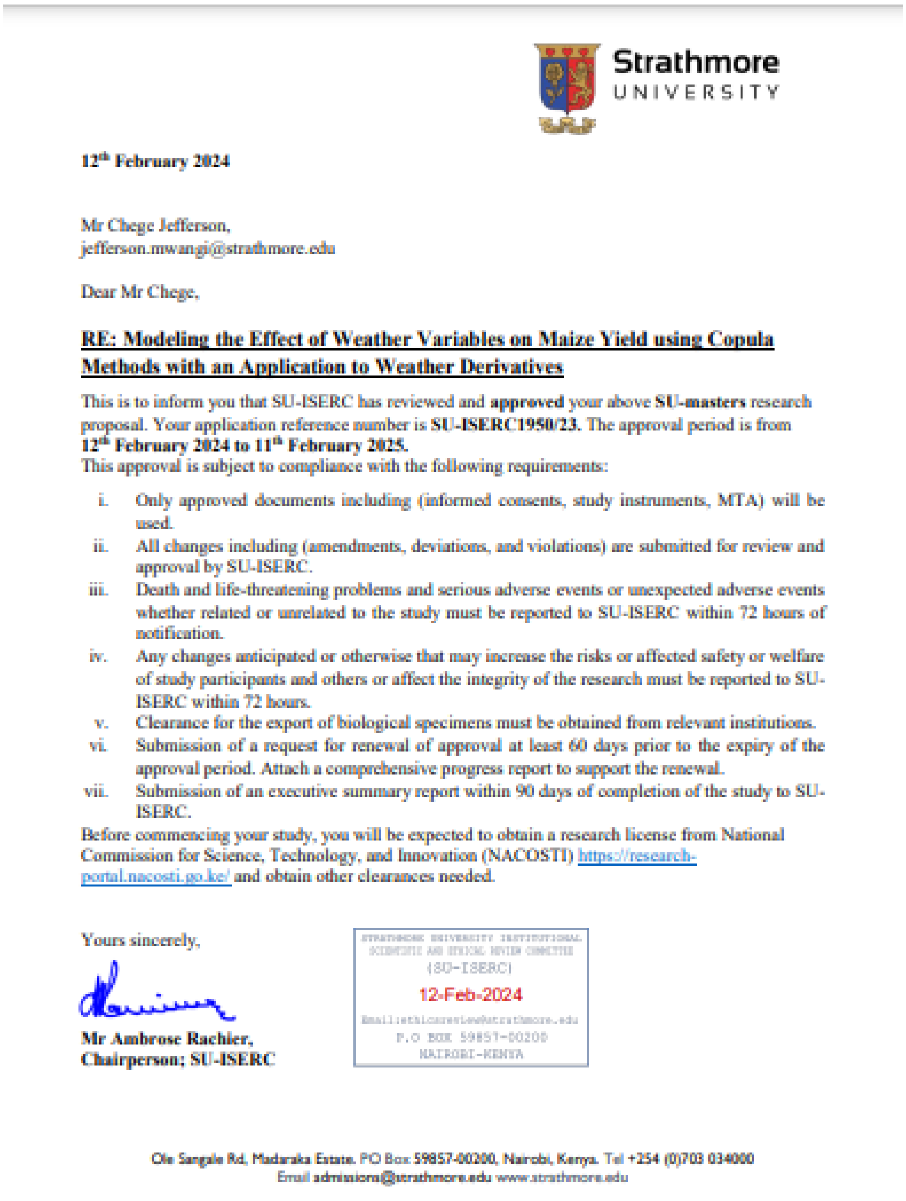
Institute of Mathematical Sciences
Strathmore University
Nairobi, Kenya

March 2025

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Figure B.1: Similarity report page 2

Chapter C: Ethical Clearance Confirmation



Strathmore
UNIVERSITY

12th February 2024

Mr Chege Jefferson,
jefferson.mwangi@strathmore.edu

Dear Mr Chege,

RE: Modeling the Effect of Weather Variables on Maize Yield using Copula Methods with an Application to Weather Derivatives

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** research proposal. Your application reference number is **SU-ISERC1950/23.** The approval period is from **12th February 2024 to 11th February 2025.**
This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used.
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 72 hours of notification.
iv. Any changes anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 72 hours.
v. Clearance for the export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

Before commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

Mr Ambrose Rachier,
Chairperson; SU-ISERC

STRATHMORE UNIVERSITY INSTITUTIONAL
SCIENTIFIC AND ETHICAL REVIEW COMMITTEE
(SU-ISERC)
12-Feb-2024
Email:ethicsreview@strathmore.edu
P.O BOX 59857-00200
NAIROBI-KENYA

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu

Figure C.1: Ethical Clearance Confirmation Letter

Chapter D: Research License from NACOSTI

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR
SCIENCE,TECHNOLOGY & INNOVATION

Ref No: 672177

Date of Issue: 15/February/2024

RESEARCH LICENSE

This is to Certify that Mr.. Jefferson Mwangi Chege of Strathmore University, has been licensed to conduct research as per the provision of the Science, Technology and Innovation Act, 2013 (Rev.2014) in on the topic: Modelling the effect of weather variables on maize yield using copula methods with an application to weather derivatives for the period ending : 15/February/2025.

License No: **NACOSTI/P/24/33208**

672177

Applicant Identification Number

Director General
NATIONAL COMMISSION FOR
SCIENCE,TECHNOLOGY &
INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document,
Scan the QR Code using QR scanner application.

See overleaf for conditions

Figure D.1: NACOSTI Clearance Confirmation Letter

Chapter E: R Code

The R code used for copula fitting and simulations in Chapter 4.

```
##Importing maize yield data
Maize_yield_data <- read_excel("C:/Users/Jefferson Chege/OneDrive/Msc Thesis
/Data analysis/Datasets/Maize yield data.xlsx")
View(Maize_yield_data)

##Importing hydro climatic data
Temp_Precpt_data <- read_excel("C:/Users/Jefferson Chege/OneDrive/Msc Thesis
/Data analysis/Datasets/Hydroclimatic data.xlsx")
View(Temp_Precpt_data)

##Defining variables
prd <- Maize_yield_data$'Production'

Temp <- Maize_yield_data$'Tmp_Mn'

precipitation <- Maize_yield_data$'Prc'

comb.df <- cbind(prd, Temp, precipitation)

comb_df <- cbind(detrended_prd, Temp, precipitation)

detrended_prd <- as.numeric(detrend(prd))

productivity <- Maize_yield_data$'Yield(Tons/Ha)'
```

```
arable_land <- Maize_yield_data$’Area(Ha per 000)’

##Summary statistics for the three variables
describe(prd)

describe(Temp)

describe(precipitation)

describe(productivity)

describe(arable_land)

##Historical time plots for temperature
historical = xts(Temp_Precpt_data[,c("Prc","Tmp_min","Tmp_max","Tmp_Mn")],
                 order.by = as.Date(Temp_Precpt_data$’Month/Year’)
                 )

historical.ts = ts_regular(historical)

historical.mod = na.fill(historical.ts,"extend")

window(historical.mod,start=as.Date("1990-01-01"),
                    end=as.Date("2020-12-31"))

plot(ts_ts(historical.mod$Tmp_max), col= ’darkred’, bty = "n", las = 1,
     fg = NA, ylim = c(10,30),
     xlab = "Time in years",
     ylab = "Temperature in degrees celsius",
     main = "Monthly Maximum & Minimum Temperature in Uasin Gishu
     between 1990 to 2020")
```

```
lines(ts_ts(historical.mod$Tmp_min), col = 'navy')

grid(col = 'gray')

legend ("topright", fill = c('darkred','navy'), cex = 0.7,
        legend = c("Tmp_max","Tmp_min"), bg = 'white')


#Historical time plots for precipitation
barplot(historical$Prc,
        col = 'black',
        space = 0,bty = "n", las = 1, fg = NA,
        xlab = "Time in years",
        ylab = "Precipitation in mm",
        main = "Monthly Precipitation in Uasin Gishu between 1990 to 2020" )


#correlation for all variables: original
pairs.panels(comb.df, smooth = TRUE, scale = FALSE,
             density=TRUE,ellipses=TRUE,
             digits = 2,method="pearson",
             hist.col="grey",show.points=TRUE)

pairs.panels(comb.df, smooth = TRUE, scale = FALSE,
             density=TRUE,ellipses=TRUE,
             digits = 2,method="spearman",
             hist.col="grey",show.points=TRUE)

pairs.panels(comb.df, smooth = TRUE, scale = FALSE,
             density=TRUE,ellipses=TRUE,
             digits = 2,method="kendall",
             hist.col="grey",show.points=TRUE)
```

```
#correlation for all variables: transformed
pairs.panels(comb_df, smooth = TRUE, scale = FALSE,
             density=TRUE,ellipses=TRUE,
             digits = 2,method="pearson",
             hist.col="grey",show.points=TRUE)

pairs.panels(comb_df, smooth = TRUE, scale = FALSE,
             density=TRUE,ellipses=TRUE,
             digits = 2,method="spearman",
             hist.col="grey",show.points=TRUE)

pairs.panels(comb_df, smooth = TRUE, scale = FALSE,
             density=TRUE,ellipses=TRUE,
             digits = 2,method="kendall",
             hist.col="grey",show.points=TRUE)

## testing for correlation: all variables (original and transformed)
Box.test(prd,type = "Ljung-Box")
Box.test(prd^2, type = "Ljung-Box")

Box.test(detrended_prd,type = "Ljung-Box")
Box.test(detrended_prd^2, type = "Ljung-Box")

Box.test(Temp, type = "Ljung-Box")
Box.test(Temp^2, type = "Ljung-Box")

Box.test(precipitation, type = "Ljung-Box")
Box.test(precipitation^2, type = "Ljung-Box")


##testing for autocorrelation in annual maize production time series
```

```
acf(
  prd,
  lag.max = NULL,
  type = c("correlation", "covariance", "partial"),
  plot = TRUE,
  main = "Series of Maize production",
  na.action = na.fail,
  demean = TRUE,
)

pacf(
  prd,
  plot = TRUE,
  main = "Series of maize production",
  na.action = na.fail,
)

##testing for autocorrelation in squared annual maize production time series
acf(
  prd^2,
  lag.max = NULL,
  type = c("correlation", "covariance", "partial"),
  plot = TRUE,
  main = "Series of squared maize production",
  na.action = na.fail,
  demean = TRUE,
)

pacf(
  prd^2,
  plot = TRUE,
  main = "Series of squared maize production",
```

```
  na.action = na.fail,
)

##testing for autocorrelation in detrended annual maize production time series
acf(
  detrended_prd,
  lag.max = NULL,
  type = c("correlation", "covariance", "partial"),
  plot = TRUE,
  main = "Series of detrended annual maize production",
  na.action = na.fail,
  demean = TRUE,
)

pacf(
  detrended_prd,
  plot = TRUE,
  main = "Series of detrended annual maize production",
  na.action = na.fail,
)

##testing for autocorrelation in squared detrended annual maize production time seri
acf(
 detrended_prd^2,
  lag.max = NULL,
  type = c("correlation", "covariance", "partial"),
  plot = TRUE,
  main = "Series of squared detrended annual maize production",
  na.action = na.fail,
  demean = TRUE,
)
```

```
pacf(
  detrended_prd^2,
  plot = TRUE,
  main = "Series of squared detrended annual maize production",
  na.action = na.fail,
)

##testing for autocorrelation in annual mean temperature time series
acf(
  Temp,
  lag.max = NULL,
  type = c("correlation", "covariance", "partial"),
  plot = TRUE,
  main = "Series of annual mean temperature",
  na.action = na.fail,
  demean = TRUE
)



acf(
  Temp,
  lag.max = NULL,
  type = c("correlation", "covariance", "partial"),
  plot = FALSE,
  na.action = na.fail,
  main = "Series of temperature",
  demean = TRUE
)

pacf(
  Temp,  main = "Series of annual mean temperature",)
```

```
##testing for autocorrelation in squared annual mean temperature time series
acf(
  Temp^2,
  lag.max = NULL,
  type = c("correlation", "covariance", "partial"),
  plot = TRUE,
  main = "Series of squared annual mean temperature",
  na.action = na.fail,
  demean = TRUE
)


acf(
  Temp^2,
  lag.max = NULL,
  type = c("correlation", "covariance", "partial"),
  plot = FALSE,
  na.action = na.fail,
  main = "Series of temperature",
  demean = TRUE
)

pacf(
  Temp^2,  main = "Series of squared annual mean temperature",)

##testing for autocorrelation in annual mean precipitation time series
acf(
  precipitation,
  lag.max = NULL,
  type = c("correlation", "covariance", "partial"),
  plot = TRUE,
  na.action = na.fail,
```

```
  main = "Series of annual mean precipitation",
  demean = TRUE
)



acf(
  precipitation,
  lag.max = NULL,
  type = c("correlation", "covariance", "partial"),
  plot = FALSE,
  na.action = na.fail,
  main = "Series of annual mean precipitation",
  demean = TRUE
)

pacf(
  precipitation,
  main = "Series of annual mean precipitation",
)

##testing for autocorrelation in squared annual mean precipitation time series
acf(
  precipitation^2,
  lag.max = NULL,
  type = c("correlation", "covariance", "partial"),
  plot = TRUE,
  na.action = na.fail,
  main = "Series of squared annual mean precipitation",
  demean = TRUE
)
```

```
acf(
  precipitation^2,
  lag.max = NULL,
  type = c("correlation", "covariance", "partial"),
  plot = FALSE,
  na.action = na.fail,
  main = "Series of squared annual mean precipitation",
  demean = TRUE
)

pacf(
  precipitation^2,
  main = "Series of squared annual mean precipitation",
)

##Testing for stationarity
kpss_test_maize_raw<- ur.kpss(prd, type = "tau")
summary(kpss_test_maize_raw)

pp_test_maize_raw<- pp.test(prd)
print(pp_test_maize_raw)

adf_test_maize_raw<- ur.df(prd, type = "trend")
summary(adf_test_maize_raw)

kpss_test_maize_det<- ur.kpss(detrended_prd, type = "tau")
summary(kpss_test_maize_det)

pp_test_maize_det<- pp.test(detrended_prd)
print(pp_test_maize_det)

adf_test_maize_det<- ur.df(detrended_prd, type = "trend")
```

```
summary(adf_test_maize_det)

kpss_test_temp<- ur.kpss(Temp, type = "tau")
summary(kpss_test_temp)

pp_test_temp<- pp.test(Temp)
print(pp_test_temp)

adf_test_temp<- ur.df(Temp, type = "trend", lags = 4)
summary(adf_test_temp)

kpss_test_prec<- ur.kpss(precipitation, type = "tau", lags = "nil")
summary(kpss_test_prec)

pp_test_prec<- pp.test(precipitation)
print(pp_test_prec)

adf_test_prec<- ur.df(precipitation, type = "trend", lags = 1)
summary(adf_test_prec)

##Density plots
ggdensity(prd,
          main = "Density plot of annual maize production between 1990 to 2020",
          xlab = " Mean annual maize production in tonnes per 000 ")

ggdensity(detrended_prd,
 main = "Density plot of detrended maize production in Uasin Gishu between 1990 to 2
          xlab = "Detrended mean annual maize production in tonnes per 000 ")

ggdensity(Temp,
main = "Density plot of temperature in Uasin Gishu between 1990 to 2020",
          xlab = "Mean annual temperature in degrees celsius")
```

```
ggdensity(precipitation,
   main = "Density plot of precipitation in Uasin Gishu between 1990 to 2020",
          xlab = "Mean annual precipitation in mm ")

##Tests for skewness and kurtosis
moments::skewness(prd)
moments::skewness(detrended_prd)
moments::skewness(Temp)
moments::skewness(precipitation)

moments::kurtosis(prd)
moments::kurtosis(detrended_prd)
moments::kurtosis(Temp)
moments::kurtosis(precipitation)

##Normality tests
shapiro.test(prd)

shapiro.test(detrended_prd)

shapiro.test(Temp)

shapiro.test(precipitation)

##QQ plots
qqPlot(prd,  main = "Q-Q Plot of maize production")

qqPlot(detrended_prd,  main = "Q-Q Plot of Detrended maize production")

qqPlot(Temp, main = "Q-Q Plot of Temperature")
```

```
qqPlot(precipitation, main = "Q-Q Plot of Precipitation")

# Fitting different distributions - precipitation
prec_norm.f <- fitdist(precipitation, "norm")
qqcomp(list(prec_norm.f),
       main = "Fitting rainfall dataset to normal distribution",
       legendtext = plot.legend)

prec_exp.f <- fitdist(precipitation, "exp")
qqcomp(list(prec_exp.f),
       main = "Fitting rainfall dataset to exponential distribution",
       legendtext = "Exponential")

df <- data.frame(precipitation)
Prec2_modified <- df[apply(df!=0, 1, all),]
Prec2_modified
prec_lnorm.f <- fitdist(Prec2_modified, "lnorm")
qqcomp(list(prec_lnorm.f),
       main = "Fitting rainfall dataset to log normal distribution",
       legendtext = "Log-normal")

Prec2_scaled <- (precipitation - min(precipitation) +
                   0.001) / (max(precipitation) - min(precipitation) + 0.002)
prec_gamma.f <- fitdist(Prec2_scaled, "gamma")
qqcomp(list(prec_gamma.f),
       main = "Fitting rainfall dataset to gamma distribution",
       legendtext = "Gamma")

# Fitting different distributions - temperature
temp_norm.f <- fitdist(Temp, "norm")
qqcomp(list(temp_norm.f),
       main = "Fitting temperature dataset to normal distribution",
```

```
       legendtext = " ")

temp_exp.f <- fitdist(Temp, "exp")
qqcomp(list(temp_exp.f),
       main = "Fitting temperature dataset to exponential distribution",
       legendtext = "exponential")

df <- data.frame(Temp)
Temp_modified <- df[apply(df!=0, 1, all),]
Temp_modified
temp_lnorm.f <- fitdist(Temp_modified, "lnorm")
qqcomp(list(temp_lnorm.f),
       main = "Fitting temperature dataset to log normal distribution",
       legendtext = "Log-normal")

Temp_scaled <- (Temp - min(Temp) +
                   0.001) / (max(Temp) - min(Temp) + 0.002)
temp_gamma.f <- fitdist(Temp_scaled, "gamma")
qqcomp(list(temp_gamma.f),
       main = "Fitting temperature dataset to gamma distribution",
       legendtext = "Gamma")

# Fitting different distributions - maize production
prd_norm.f <- fitdist(detrended_prd, "norm")
qqcomp(list(prd_norm.f),
       main = "Fitting normal distribution to detrended
       maize production dataset",
       legendtext = plot.legend)

prd_exp.f <- fitdist(prd_scaled, "exp")
qqcomp(list(prd_exp.f),
       main = "Fitting exponential distribution to detrended
```

```
        maize production dataset",
        legendtext = "Exponential")

prd_lnorm.f <- fitdist(prd_scaled, "lnorm")
qqcomp(list(prd_lnorm.f),
        main = "Fitting log normal distribution to detrended
        maize production dataset",
        legendtext = "Log-normal")

prd_scaled <- (detrended_prd - min(detrended_prd) +
                  0.001) / (max(detrended_prd) - min(detrended_prd) + 0.001)

prd_gamma.f <- fitdist(prd_scaled, "gamma")
qqcomp(list(prd_gamma.f),
        main = "Fitting gamma distribution to detrended
        maize production dataset",
        legendtext = "Gamma")

#Fitting Temperature and Precipitation copulas using MLE method
u_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 1]
v_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 2]
selectedCopula.tp.1.mle <- BiCopSelect(u_tp,
                              v_tp,
                              familyset = 1,
                              selectioncrit = "AIC",
                              indeptest = FALSE,
                              presel = TRUE,
                              method = "mle")
summary(selectedCopula.tp.1.mle)

u_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 1]
v_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 2]
```

```
selectedCopula.tp.2.mle <- BiCopSelect(u_tp,
                             v_tp,
                             familyset = 2,
                             selectioncrit = "AIC",
                             method = "mle")
summary(selectedCopula.tp.2.mle)

u_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 1]
v_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 2]
selectedCopula.tp.3.mle <- BiCopSelect(u_tp,
                             v_tp,
                             familyset = 3,
                             selectioncrit = "AIC",
                             method = "mle")
summary(selectedCopula.tp.3.mle)

u_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 1]
v_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 2]
selectedCopula.tp.4.mle <- BiCopSelect(u_tp,
                             v_tp,
                             familyset = 4,
                             selectioncrit = "AIC",
                             method = "mle")
summary(selectedCopula.tp.4.mle)

u_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 1]
v_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 2]
selectedCopula.tp.5.mle <- BiCopSelect(u_tp,
                             v_tp,
                             familyset = 5,
                             selectioncrit = "AIC",
                             method = "mle")
```

```
summary(selectedCopula.tp.5.mle)

#Fitting Temperature and Precipitation copulas using i tau method
u_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 1]
v_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 2]
selectedCopula.tp.1.itau <- BiCopSelect(u_tp,
                              v_tp,
                              familyset = 1,
                              selectioncrit = "AIC",
                              method = "itau")
summary(selectedCopula.tp.1.itau)

u_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 1]
v_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 2]
selectedCopula.tp.2.itau <- BiCopSelect(u_tp,
                              v_tp,
                              familyset = 2,
                              selectioncrit = "AIC",
                              method = "itau")
summary(selectedCopula.tp.2.itau)

u_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 1]
v_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 2]
selectedCopula.tp.3.itau <- BiCopSelect(u_tp,
                              v_tp,
                              familyset = 3,
                              selectioncrit = "AIC",
                              method = "itau")
summary(selectedCopula.tp.3.itau)

u_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 1]
```

```
v_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 2]
selectedCopula.tp.4.itau <- BiCopSelect(u_tp,
                              v_tp,
                              familyset = 4,
                              selectioncrit = "AIC",
                              method = "itau")
summary(selectedCopula.tp.4.itau)

u_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 1]
v_tp <- pobs(as.matrix(cbind(Temp, precipitation)))[, 2]
selectedCopula.tp.5.itau <- BiCopSelect(u_tp,
                              v_tp,
                              familyset = 5,
                              selectioncrit = "AIC",
                              method = "itau")
summary(selectedCopula.tp.5.itau)

#Fitting maize production and precipitation copulas using MLE method
u_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 1]
v_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 2]
selectedCopula.mp.1.mle <- BiCopSelect(u_mp,
                              v_mp,
                              familyset = 1,
                              selectioncrit = "AIC",
                              method = "mle")
summary(selectedCopula.mp.1.mle)

u_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 1]
v_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 2]
selectedCopula.mp.2.mle <- BiCopSelect(u_mp,
                              v_mp,
                              familyset = 2,
```

```
                              selectioncrit = "AIC",
                              method = "mle")
summary(selectedCopula.mp.2.mle)

u_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 1]
v_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 2]
selectedCopula.mp.3.mle <- BiCopSelect(u_mp,
                              v_mp,
                              familyset = 3,
                              selectioncrit = "AIC",
                              method = "mle")
summary(selectedCopula.mp.3.mle)

u_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 1]
v_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 2]
selectedCopula.mp.4.mle <- BiCopSelect(u_mp,
                              v_mp,
                              familyset = 4,
                              selectioncrit = "AIC",
                              method = "mle")
summary(selectedCopula.mp.4.mle)

u_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 1]
v_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 2]
selectedCopula.mp.5.mle <- BiCopSelect(u_mp,
                              v_mp,
                              familyset = 5,
                              selectioncrit = "AIC",
                              method = "mle")
summary(selectedCopula.mp.5.mle)

#fitting maize production and Precipitation copulas using i tau method
```

```
u_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 1]
v_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 2]
selectedCopula.mp.1.itau <- BiCopSelect(u_mp,
                              v_mp,
                              familyset = 1,
                              selectioncrit = "AIC",
                              method = "itau")
summary(selectedCopula.mp.1.itau)

u_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 1]
v_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 2]
selectedCopula.mp.2.itau <- BiCopSelect(u_mp,
                              v_mp,
                              familyset = 2,
                              selectioncrit = "AIC",
                              method = "itau")
summary(selectedCopula.mp.2.itau)

u_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 1]
v_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 2]
selectedCopula.mp.3.itau <- BiCopSelect(u_mp,
                              v_mp,
                              familyset = 3,
                              selectioncrit = "AIC",
                              method = "itau")
summary(selectedCopula.mp.3.itau)

u_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 1]
v_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 2]
selectedCopula.mp.4.itau <- BiCopSelect(u_mp,
                              v_mp,
                              familyset = 4,
```

```
                              selectioncrit = "AIC",
                              method = "itau")
summary(selectedCopula.mp.4.itau)

u_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 1]
v_mp <- pobs(as.matrix(cbind(detrended_prd, precipitation)))[, 2]
selectedCopula.mp.5.itau <- BiCopSelect(u_mp,
                              v_mp,
                              familyset = 5,
                              selectioncrit = "AIC",
                              method = "itau")
summary(selectedCopula.mp.5.itau)



#fitting maize production and temperature copulas using MLE method
u_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 1]
v_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 2]
selectedCopula.mt.1.mle <- BiCopSelect(u_mt,
                              v_mt,
                              familyset = 1,
                              selectioncrit = "AIC",
                              method = "mle")
summary(selectedCopula.mt.1.mle)

u_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 1]
v_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 2]
selectedCopula.mt.2.mle <- BiCopSelect(u_mt,
                              v_mt,
                              familyset = 2,
                              selectioncrit = "AIC",
                              method = "mle")
summary(selectedCopula.mt.2.mle)
```

```
u_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 1]
v_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 2]
selectedCopula.mt.3.mle <- BiCopSelect(u_mt,
                              v_mt,
                              familyset = 3,
                              selectioncrit = "AIC",
                              method = "mle")
summary(selectedCopula.mt.3.mle)

u_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 1]
v_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 2]
selectedCopula.mt.4.mle <- BiCopSelect(u_mt,
                              v_mt,
                              familyset = 4,
                              selectioncrit = "AIC",
                              method = "mle")
summary(selectedCopula.mt.4.mle)

u_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 1]
v_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 2]
selectedCopula.mt.5.mle <- BiCopSelect(u_mt,
                              v_mt,
                              familyset = 5,
                              selectioncrit = "AIC",
                              method = "mle")
summary(selectedCopula.mt.5.mle)

#fitting maize production and Temperature copulas using i tau method
u_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 1]
v_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 2]
selectedCopula.mt.1.itau <- BiCopSelect(u_mt,
```

```
                              v_mt,
                              familyset = 1,
                              selectioncrit = "AIC",
                              method = "itau")
summary(selectedCopula.mt.1.itau)

u_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 1]
v_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 2]
selectedCopula.mt.2.itau <- BiCopSelect(u_mt,
                              v_mt,
                              familyset = 2,
                              selectioncrit = "AIC",
                              method = "itau")
summary(selectedCopula.mt.2.itau)

u_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 1]
v_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 2]
selectedCopula.mt.3.itau <- BiCopSelect(u_mt,
                              v_mt,
                              familyset = 3,
                              selectioncrit = "AIC",
                              method = "itau")
summary(selectedCopula.mt.3.itau)

u_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 1]
v_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 2]
selectedCopula.mt.4.itau <- BiCopSelect(u_mt,
                              v_mt,
                              familyset = 4,
                              selectioncrit = "AIC",
                              method = "itau")
summary(selectedCopula.mt.4.itau)
```

```
u_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 1]
v_mt <- pobs(as.matrix(cbind(detrended_prd, Temp)))[, 2]
selectedCopula.mt.5.itau <- BiCopSelect(u_mt,
                              v_mt,
                              familyset = 5,
                              selectioncrit = "AIC"
                              ,
                              method = "itau")
summary(selectedCopula.mt.5.itau)


#Trivariate Copula fitting
Copdat1 <- matrix(c(detrended_prd, precipitation,
                   Temp), nrow = 31, ncol = 3)
u_cop<- pobs(Copdat1)
u.cop<-as.matrix(u_cop)

norm.cop<- copula::fitCopula(normalCopula(dim = 3, dispstr = "un"),
                    u.cop, method = "ml")
summary(norm.cop)

AIC_norm<- (-2)*(norm.cop@loglik)+2*(length(norm.cop@estimate))
AIC_norm
BIC_norm<-  (-2)*(norm.cop@loglik)+log(31)*(length(norm.cop@estimate))
BIC_norm

norm.cop_itau<- copula::fitCopula(normalCopula(dim = 3, dispstr = "un"),
                             u.cop, method = "itau")
summary(norm.cop_itau)
```

```
t.cop_ml<- copula::fitCopula(tCopula(dim = 3, dispstr = "un"),
                         u.cop,method = "ml")
summary(t.cop_ml)

t.cop_ml@loglik

AIC_t_ml<- (-2)*(t.cop_ml@loglik)+2*(length(t.cop_ml@estimate))
AIC_t_ml
BIC_t_ml<-  (-2)*(t.cop_ml@loglik)+log(31)*(length(t.cop_ml@estimate))
BIC_t_ml

t.cop_itau<- copula::fitCopula(tCopula(dim = 3, dispstr = "un"),
                         u.cop,method = "itau")
summary(t.cop_itau)

clay.cop <- copula::fitCopula(claytonCopula(dim = 3),u.cop, start = 0.5,
                      method = "ml")
summary(clay.cop)

clay.cop@loglik

AIC_clay <- (-2)*(clay.cop@loglik)+2*((length(clay.cop@estimate)))
AIC_clay
BIC_clay <- (-2)*(clay.cop@loglik)+log(31)*(length(clay.cop@estimate))
BIC_clay

clay.cop_itau <- copula::fitCopula(claytonCopula(dim = 3),u.cop, start = 0.5,
                              method = "itau")
summary(clay.cop_itau)

gumbel.cop <- copula::fitCopula(gumbelCopula(dim = 3),u.cop, start = 1,
                        method = "ml")
```

```
summary(gumbel.cop)

gumbel.cop@loglik

AIC_gumb <- (-2)*(gumbel.cop@loglik)+2*((length(gumbel.cop@estimate)))
AIC_gumb
BIC_gumb <- (-2)*(gumbel.cop@loglik)+log(31)*(length(gumbel.cop@estimate))
BIC_gumb

gumbel.cop_itau <- copula::fitCopula(gumbelCopula(dim = 3),u.cop, start = 1,
                                method = "itau")
summary(gumbel.cop_itau)

frank.cop <- copula::fitCopula(frankCopula(dim = 3),u.cop, start = 0.5,
                      method = "ml")
summary(frank.cop)

frank.cop@loglik

AIC_frank <- (-2)*(frank.cop@loglik)+2*((length(frank.cop@estimate)))
AIC_frank
BIC_frank <- (-2)*(frank.cop@loglik)+log(31)*(length(frank.cop@estimate))
BIC_frank

frank.cop_itau <- copula::fitCopula(frankCopula(dim = 3),u.cop, start = 0.5,
                               method = "itau")
summary(frank.cop_itau)

#Fitting conditional copulas using conditional distribution functions

#Defining the marginals
U1<-pobs(Copdat1)
```

```
u1<-U1[,1]
u2<-U1[,2]
u3<-U1[,3]

# Extracting the first pair-copula for V1|V2
pair_copula_1_2 <- BiCopSelect(u1, u2, familyset = c(1,2,3,4,5))
summary(pair_copula_1_2)

# Calculate conditional distribution of  V1|V2
h_V1_given_V2 <- BiCopHfunc(u1, u2, pair_copula_1_2)
h_V1_given_V2

# Extracting the second pair-copula for V1|V3
pair_copula_1_3 <- BiCopSelect(u1, u3, familyset = c(1,2,3,4,5))
summary(pair_copula_1_3)

# Calculate conditional distribution of  V1|V3
h_V1_given_V3 <- BiCopHfunc(u1, u3, pair_copula_1_3)
h_V1_given_V3

# Extracting the third pair-copula for V2|V3
pair_copula_2_3 <- BiCopSelect(u2, u3, familyset = c(1,2,3,4,5))
summary(pair_copula_2_3)

# Calculate conditional distribution of  V2|V3
h_V2_given_V3 <- BiCopHfunc(u2, u3, pair_copula_2_3)
h_V2_given_V3

#Estimate the conditional copula for 1,3 given 2
one_three_given_2.mle<-BiCopSelect(h_V1_given_V2$hfunc1,
                            h_V2_given_V3$hfunc1, familyset = NA,
                            method = "mle")
```

```
summary(one_three_given_2.mle)

one_three_given_2.itau<-BiCopSelect(h_V1_given_V2$hfunc1,
                                   h_V2_given_V3$hfunc1, familyset = NA,
                                   method = "itau")

summary(one_three_given_2.itau)

#Estimate the conditional copula for 2,3 given 1
two_three_given_1.mle<-BiCopSelect(h_V1_given_V2$hfunc1,
                               h_V1_given_V3$hfunc1,
                            familyset = NA, method = "mle")
summary(two_three_given_1.mle)

two_three_given_1.itau<-BiCopSelect(h_V1_given_V2$hfunc1,
                                   h_V1_given_V3$hfunc1,
                                   familyset = NA, method = "itau")
summary(two_three_given_1.itau)

#Estimate the conditional copula for 1,2 given 3
one_two_given_3.mle<-BiCopSelect(h_V1_given_V3$hfunc1,
                            h_V2_given_V3$hfunc1, familyset = NA,
                            method = "mle")
summary(one_two_given_3.mle)

one_two_given_3.itau<-BiCopSelect(h_V1_given_V3$hfunc1,
                                 h_V2_given_V3$hfunc1, familyset = NA,
                                 method = "itau")
summary(one_two_given_3.itau)

#Fitting Vine Copulas - 1,3|2
```

```
# Set random seed for reproducibility
set.seed(123)

# Define the TRUE vine copula (the assumed model)
pclist <- list(
  list( bicop_dist(’frank’, parameters=1.66),
        bicop_dist(’gumbel’,rotation = 270, parameters=1.67)),
  list(bicop_dist(’indep’)))


vmat<-matrix(c(2,3,1,
               3,2,0,
               3,0,0), ncol = 3)

vstruct<- as_rvine_structure(vmat)

vcop_true <- vinecop_dist(pclist, vstruct)

summary(vcop_true)

contour(vcop_true)

# Generate samples and transform to std. normal margins.
z <- qnorm(rvinecop(10000, vcop_true))
colnames(z) <- paste0(’z’, seq(1,3))

# Visualize normal distribution charts for generated random samples
zf <- as.data.frame(z)
colnames(zf) <- paste0(’z’, 1:3)
GGally::ggpairs(zf, progress=FALSE)
```

```
# Pseudo-observations for copula fitting
u <- pseudo_obs(z)
colnames(u) <- paste0('u', 1:3)

#Fit a NEW vine copula to the data
vinefit1_non.arch <- vinecop(u,family_set =  c("t","gaussian"
                                        ), ,
                             structure = vstruct)
vinefit1_non.arch
summary(vinefit1_non.arch)
contour(vinefit1_non.arch)

vinefit1_arch <- vinecop(u,family_set =  c("clayton",
                                        "gumbel", "frank"), ,
                         structure = vstruct)


vinefit1_arch
summary(vinefit1_arch)
contour(vinefit1_arch)

vinefit1_nonparam <- vinecop(u, family_set = c('tll'), ,
                             structure = vstruct)
vinefit1_nonparam
summary(vinefit1_nonparam)
contour(vinefit1_nonparam)

#Fitting Vine Copulas - 1,2|3

# Set random seed for reproducibility
set.seed(123)

# Define the TRUE vine copula (the assumed model)
```

```
pclist2 <- list(
  list(bicop_dist(’gaussian’, parameters=-0.18),
       bicop_dist(’gumbel’, rotation = 270, parameters=1.67)),
  list(bicop_dist(’gaussian’, parameters=0.84)))

vmat2<-matrix(c(3,2,1,
                2,3,0,
                2,0,0), ncol = 3)

vstruct2<- as_rvine_structure(vmat2)

vcop_true2 <- vinecop_dist(pclist2, vstruct2)

summary(vcop_true2)

contour(vcop_true2)

## Generate samples and transform to std. normal margins.
z2 <- qnorm(rvinecop(10000, vcop_true2))
colnames(z2) <- paste0(’z’, seq(1,3))

# Visualize normal distribution charts for generated random samples
zf2 <- as.data.frame(z2)
colnames(zf2) <- paste0(’z’, seq(1,3))
GGally::ggpairs(zf2, progress=FALSE)

#Pseudo-observations for copula fitting
u4 <- pseudo_obs(z2)
colnames(u4) <- paste0(’u’, seq(1,3))

#Fit a NEW vine copula to the data
vinefit2_non.arch <- vinecop(u4,family_set = c("t","gaussian"
```

```
                                               ),
                             structure = vstruct2)
vinefit2_non.arch
summary(vinefit2_non.arch)
contour(vinefit2_non.arch)

vinefit2_arch <- vinecop(u4,family_set = c("clayton",
                                           "gumbel", "frank"),
                         structure = vstruct2)
vinefit2_arch
summary(vinefit2_arch)
contour(vinefit2_arch)

vinefit2_nonparam <- vinecop(u4, family_set = c(’tll’),
                             structure = vstruct2)
vinefit2_nonparam
summary(vinefit2_nonparam)
contour(vinefit2_nonparam)


#Fitting Vine Copulas - 2,3|1

# Set random seed for reproducibility
set.seed(123)

# Define the TRUE vine copula (the assumed model)
pclist3 <- list(
  list( bicop_dist(’frank’,parameters=1.66),
        bicop_dist(’gaussian’, parameters=-0.18)),
  list(bicop_dist(’gumbel’, rotation = 90, parameters=1.64)))
```

```
vmat3<-matrix(c(1,3,2,
                3,1,0,
                3,0,0), ncol = 3)


vstruct3<- as_rvine_structure(vmat3)

vcop_true3 <- vinecop_dist(pclist3, vstruct3)

summary(vcop_true3)

contour(vcop_true3)

## Generate samples and transform to std. normal margins.
z3 <- qnorm(rvinecop(10000, vcop_true3))
colnames(z3) <- paste0('z', seq(1,3))

# Visualize normal distribution charts for generated random samples
zf3 <- as.data.frame(z3)
colnames(zf3) <- paste0('z', seq(1,3))
GGally::ggpairs(zf3, progress=FALSE)

# Pseudo-observations for copula fitting
u5 <- pseudo_obs(z3)
colnames(u5) <- paste0('u', seq(1,3))

#Fit a NEW vine copula to the data
vinefit3_non.arch <- vinecop(u5,family_set = c("t","gaussian"
                                               ),
                      structure = vstruct3)
vinefit3_non.arch
summary(vinefit3_non.arch)
```

```
contour(vinefit3_non.arch)


vinefit3_arch <- vinecop(u5,family_set = c("clayton",
                                           "gumbel", "frank"),
                             structure = vstruct3)
vinefit3_arch
summary(vinefit3_arch)
contour(vinefit3_arch)



vinefit3_nonparam <- vinecop(u5, family_set = c('tll'),
                             structure = vstruct3)
vinefit3_nonparam
summary(vinefit3_nonparam)
contour(vinefit3_nonparam)
```